

Epigenetic clock analysis of diet, exercise, education, and lifestyle factors

Austin Quach^{1*}, Morgan E. Levine^{1*}, Toshiko Tanaka^{2*}, Ake T. Lu¹, Brian H. Chen², Luigi Ferrucci², Beate Ritz^{3,4}, Stefania Bandinelli⁵, Marian L. Neuhauser⁶, Jeannette M. Beasley⁷, Linda Snetselaar⁸, Robert B. Wallace⁸, Philip S. Tsao^{9,10}, Devin Absher¹¹, Themistocles L. Assimes⁹, James D. Stewart¹², Yun Li^{13,14}, Lifang Hou^{15,16}, Andrea A. Baccarelli¹⁷, Eric A. Whitset^{12,18}, Steve Horvath^{1,19}

¹Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA

²Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, National Institutes of Health, USA. Baltimore, MD 21224, USA

³Department of Neurology, UCLA School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA

⁴Department of Epidemiology, UCLA Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA 90095, USA

⁵Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy

⁶Cancer Prevention Program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

⁷Department of Medicine, New York University, New York, NY 10016, USA

⁸Department of Epidemiology, University of Iowa, 145 N. Riverside Drive, Iowa City, IA 52242, USA

⁹Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

¹⁰VA Palo Alto Health Care System, Palo Alto CA 94304, USA

¹¹HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA

¹²Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599, USA

¹³Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA

¹⁴Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599, USA

¹⁵Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University Chicago, IL 60611, USA

¹⁶Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University Chicago, IL 60611, USA

¹⁷Laboratory of Environmental Epigenetics, Departments of Environmental Health Sciences Epidemiology, Columbia University Mailman School of Public Health, New York, NY 10032, USA

¹⁸Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA

¹⁹Dept. of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA 90095, USA

Correspondence to: Steve Horvath; **email:** shorvath@mednet.ucla.edu

Keywords: diet, lifestyle, fish intake, alcohol intake, aging, epigenetic clock, DNA methylation

Received: November 11, 2016

Accepted: January 25, 2017

Published: February 14, 2017

ABSTRACT

Behavioral and lifestyle factors have been shown to relate to a number of health-related outcomes, yet there is a need for studies that examine their relationship to molecular aging rates. Toward this end, we use recent epi-

genetic biomarkers of age that have previously been shown to predict all-cause mortality, chronic conditions and age-related functional decline. We analyze cross-sectional data from 4,173 postmenopausal female participants from the Women's Health Initiative, as well as 402 male and female participants from the Italian cohort study, *Invecchiare nel Chianti*.

Extrinsic epigenetic age acceleration (EEAA) exhibits significant associations with fish intake ($p=0.02$), moderate alcohol consumption ($p=0.01$), education ($p=3 \times 10^{-5}$), BMI ($p=0.01$), and blood carotenoid levels ($p=1 \times 10^{-5}$)—an indicator of fruit and vegetable consumption, whereas intrinsic epigenetic age acceleration (IEAA) is associated with poultry intake ($p=0.03$) and BMI ($p=0.05$). Both EEAA and IEAA were also found to relate to indicators of metabolic syndrome, which appear to mediate their associations with BMI. Metformin—the first-line medication for the treatment of type 2 diabetes—does not delay epigenetic aging in this observational study. Finally, longitudinal data suggests that an increase in BMI is associated with increase in both EEAA and IEAA.

Overall, the epigenetic age analysis of blood confirms the conventional wisdom regarding the benefits of eating a high plant diet with lean meats, moderate alcohol consumption, physical activity, and education, as well as the health risks of obesity and metabolic syndrome.

INTRODUCTION

A number of behavioral lifestyle factors have been shown to relate to health, including diet, physical activity, moderate alcohol consumption, and educational attainment. For instance, diet is a modifiable behavior with the potential to mitigate chronic disease risk. Various dietary components have been reported to influence intermediate risk factors and the prevalence of age-related disease outcomes; thus there is a growing consensus regarding nutritional recommendations for maintaining optimal health. These dietary factors include whole grain & dietary fiber [1], fish & omega-3 fatty acids [2], and fruits & vegetables [3], all of which may be involved in reducing systemic inflammation [4]. Further, metabolic health has been established as one of the primary mechanisms through which diet affects health and disease [5]. Conditions such as, insulin resistance, hypercholesterolemia, hypertension, hypertriglyceremia, and systemic inflammation can be promoted by poor dietary habits and often coalesce, influencing a person's risk of atherosclerosis, diabetes mellitus, and stroke [6-8].

In addition to diet, other behaviors such as moderate alcohol consumption, increased physical activity, and higher educational attainment have all been linked to reductions in morbidity and mortality risk [9-16]. Yet, despite the strong evidence connecting lifestyle factors to health outcomes, it is still unclear whether these factors directly influence aging on a molecular level. In previous work, leukocyte telomere length (LTL) has been used to investigate the influence of lifestyle factors on replicative aging in blood [17-21]. A cross-sectional study of 2,284 participants from the Nurses' Health Study reported that LTL was associated with BMI, waist circumference, and dietary intake of total fat, polyunsaturated fatty acids, and fiber [22]. LTL was also found to be longer among individuals who were

more physically active [23, 24], as well as those with higher levels of education [25].

Another promising measure for investigating the dynamics between lifestyle and aging is the molecular biomarker known as the "epigenetic clock". Chronological age has been shown to have a profound effect on DNA methylation levels [26-34]. As a result, several highly accurate epigenetic biomarkers of chronological age have been proposed [35-39]. These biomarkers use weighted averages of methylation levels at specific CpG sites to produce estimates of age (in units of years), referred to as "DNA methylation age" (DNAm age) or "epigenetic age". Recent studies support the idea that these measures are at least passive biomarkers of biological age. For instance, the epigenetic age of blood has been found to be predictive of all-cause mortality [40-43], frailty [44], lung cancer [45], and cognitive and physical functioning [46], while the blood of the offspring of Italian semi-super-centenarians (i.e. participants aged 105 or older) was shown to have a lower epigenetic age than that of age-matched controls [47]. Further, the utility of the epigenetic clock method using various tissues and organs has been demonstrated in applications surrounding Alzheimer's disease [48], centenarian status [47, 49], development [50], Down syndrome [51], frailty [44], HIV infection [52], Huntington's disease [53], obesity [54], lifetime stress [55], menopause [56], osteoarthritis [57], and Parkinson's disease [58].

However, relatively little is known about the relationship between epigenetic aging rates and lifestyle factors, such as diet, alcohol consumption, physical activity, and educational attainment. Here, we investigate these relationships by leveraging blood DNA methylation data from two large epidemiological cohorts. In our primary analysis, we use data from older women within the Women's Health Initiative (WHI) to

examine the relationships between epigenetic age acceleration in blood and dietary variables, education, alcohol, and exercise. In our secondary analysis, we sought to validate the results in the Invecchiare nel Chianti (InCHIANTI) Study, which is a population-based prospective cohort study of residents ages 21 or older from two areas in the Chianti region of Tuscany, Italy.

Since our study revealed that metabolic syndrome is associated with accelerated epigenetic aging, we also carried out a post-hoc analysis that evaluated the effect of metformin, which is a widely-used medication against type 2 diabetes.

RESULTS

Sample characteristics

The WHI sample consisted of 4,173 postmenopausal women including 2,045 Caucasians, 1,192 African Americans, and 717 Hispanics. Chronological age ranged from 50-82 years (mean=64, s.d.=7.1). The InCHIANTI sample was composed of 402 participants from a European (Italian) population, including 178 men (44%) and 229 women (56%). We used the most current cross-sectional wave for this cohort, and at that time-point participants ranged in age from 30 to 100 years (mean=71, s.d.=16). Additional details on participant characteristics can be found in the Methods and in Table 1.

Dietary and metabolic associations with measures of age acceleration

Here we leverage two distinct measures of epigenetic age acceleration which are based on different sets of CpGs: *intrinsic* epigenetic age acceleration (IEAA), and *extrinsic* epigenetic age acceleration (EEAA) (Methods). Epigenetic age acceleration is broadly defined as the epigenetic age left unexplained by chronological age, where intrinsic and extrinsic denote additional modifications to this concept. In addition to adjusting for chronological age, IEAA also adjusts the epigenetic clock for blood cell count estimates, arriving at a measure that is unaffected by both variation in chronological age and blood cell composition. EEAA, on the other hand, integrates known age-related changes in blood cell counts with a blood-based measure of epigenetic age [37] before adjusting for chronological age, making EEAA dependent on age-related changes in blood cell composition. In essence, IEAA can be interpreted as a measure of cell-intrinsic aging and EEAA as a measure of immune system aging, where for both, a positive value indicates that the epigenetic age of an individual (organ or tissue) is higher than

expected based on their chronological age—or that the individual is exhibiting accelerated epigenetic aging. We find that IEAA is only moderately correlated with EEAA ($r=0.37$), and that measurements on the same individuals at different time points (mean difference 3.0 years between visit dates) showed moderately strong correlations (IEAA $r=0.70$, EEAA $r=0.66$).

We first used a robust correlation test to relate our two measures of epigenetic aging (IEAA and EEAA) to select reported dietary exposures, blood nutrient levels, cardiometabolic plasma biomarkers, and lifestyle factors, designating a Bonferroni-corrected significance threshold of $\alpha=7 \times 10^{-4}$ (Figure 1). The correlation test results for specific racial/ethnic groups are presented in Supplementary Figure 1 and select marginal associations are shown as bar plots in Supplementary Figure 2. Pairwise correlations between dietary variables, metabolic biomarkers, and lifestyle factors are presented in Supplementary Figure 3.

EEAA exhibits weak but statistically significant correlations with fish intake ($r=-0.07$, $p=2 \times 10^{-5}$), alcohol consumption ($r=-0.07$, $p=3 \times 10^{-5}$, Supplementary Figure 4), plasma levels of mean carotenoids ($r=-0.13$, $p=2 \times 10^{-9}$), alpha-carotene ($r=-0.11$, $p=9 \times 10^{-8}$), beta-carotene ($r=-0.11$, $p=3 \times 10^{-7}$), lutein+zeaxanthin ($r=-0.9$, $p=1 \times 10^{-5}$), beta-cryptoxanthin ($r=-0.11$, $p=3 \times 10^{-7}$), gamma-tocopherol ($r=0.09$, $p=9 \times 10^{-6}$), triglyceride ($r=0.7$, $p=6 \times 10^{-6}$), C-reactive protein (CRP, $r=0.12$, $p=2 \times 10^{-10}$), insulin ($r=0.11$, $p=3 \times 10^{-12}$), HDL cholesterol ($r=-0.09$, $p=2 \times 10^{-8}$), glucose ($r=0.06$, $p=2 \times 10^{-4}$), systolic blood pressure ($r=0.07$, $p=4 \times 10^{-6}$), waist-to-hip ratio (WHR, $r=0.09$, $p=2 \times 10^{-8}$), BMI ($r=0.09$, $p=2 \times 10^{-8}$), education ($r=-0.10$, $p=3 \times 10^{-10}$), income ($r=-0.06$, $p=1 \times 10^{-4}$), and exercise ($r=-0.07$, $p=2 \times 10^{-5}$, Figure 1). In contrast, the intrinsic epigenetic aging rate exhibits weaker correlations with dietary variables and lifestyle factors: IEAA is only associated with BMI ($r=0.08$, $p=1 \times 10^{-6}$), and plasma levels of gamma-tocopherol ($r=0.08$, $p=2 \times 10^{-4}$), CRP ($r=0.08$, $p=6 \times 10^{-5}$), insulin ($r=0.07$, $p=2 \times 10^{-5}$), glucose ($r=0.06$, $p=8 \times 10^{-5}$), and triglyceride levels ($r=0.05$, $p=5 \times 10^{-4}$, Figure 1).

Meta-analysis of multivariable linear models link epigenetic age acceleration to diet

Associations with EEAA

We have recently shown that ethnicity relates to epigenetic aging rates: e.g. Hispanics have lower levels of IEAA compared to other ethnic groups [59]. Given the potential for confounding by sociodemographic and lifestyle factors, we used Stouffer's method to meta-analyze multivariate linear models, stratified by racial/ethnic group, in order to re-examine the

suggestive associations from our marginal correlation analysis. After adjusting for sex and dataset (Figure 2A), we find that lower EEAA is significantly associated with greater intake of fish ($t_{meta}=-2.92$, $p_{meta}=0.003$), higher education ($t_{meta}=-4.14$, $p_{meta}=3 \times 10^{-5}$), lower BMI ($t_{meta}=4.86$, $p_{meta}=1 \times 10^{-6}$), and current drinker status ($t_{meta}=-3.23$, $p_{meta}=0.001$). However, we

find no association for current smoking status, and poultry intake, and only a trend toward association with physical activity ($t_{meta}=-1.70$, $p_{meta}=0.09$). In the subset of WHI participants with circulating carotenoid measurements, we also find that mean carotenoid levels are associated with EEAA ($t_{meta}=-4.34$, $p_{meta}=1 \times 10^{-5}$, Supplementary Figure 5A).

		Pooled WHI samples					
		Adjusted for ethnicity and dataset					
		n	μ	IEAA		EEAA	
bicor	p			bicor	p		
Diet	log2(Total energy)	3687	10.53	0.00	0.96	-0.02	0.19
	Carbohydrate	3687	49.01	0.02	0.29	0.00	0.96
	Protein	3687	16.50	-0.02	0.15	-0.03	0.10
	Fat	3687	34.66	0.00	0.97	0.02	0.15
	log2(1+Red meat)	3687	0.75	0.03	0.10	0.02	0.28
	log2(1+Poultry)	3687	0.45	-0.05	4E-3	-0.03	0.05
	log2(1+Fish)	3687	0.31	-0.02	0.30	-0.07	2E-5
	log2(1+Dairy)	3687	1.25	0.00	0.99	-0.02	0.29
	log2(1+Whole grains)	3687	1.03	0.00	0.85	-0.02	0.19
	log2(1+Nuts)	3687	0.19	0.01	0.51	-0.02	0.36
	log2(Fruits)	3687	0.32	0.00	0.81	-0.03	0.04
	log2(Vegetables)	3687	0.62	0.00	0.98	-0.04	0.01
Blood nutrients	Retinol	2268	0.59	0.02	0.46	-0.01	0.69
	Mean carotenoids	2267	0.01	-0.06	4E-3	-0.13	2E-9
	Lycopene	2268	0.40	-0.02	0.44	-0.03	0.17
	log2(alpha-Carotene)	2268	-4.22	-0.04	0.04	-0.11	9E-8
	log2(beta-Carotene)	2267	-2.18	-0.06	0.01	-0.11	3E-7
	log2(Lutein+Zeaxanthin)	2268	-2.38	-0.04	0.09	-0.09	1E-5
	log2(beta-Cryptoxanthin)	2268	-3.74	-0.06	2E-3	-0.11	3E-7
	log2(alpha-Tocopherol)	2268	3.94	-0.04	0.07	-0.06	0.01
	log2(gamma-Tocopherol)	2268	0.68	0.08	2E-4	0.09	9E-6
Measurements	log2(C-reactive protein)	2809	1.54	0.08	6E-5	0.12	2E-10
	log2(Insulin)	4043	5.81	0.07	2E-5	0.11	3E-12
	log2(Glucose)	4145	6.66	0.06	8E-5	0.06	2E-4
	log2(Triglyceride)	4149	7.05	0.05	5E-4	0.07	6E-6
	Total cholesterol	4149	227.31	0.03	0.04	0.01	0.62
	LDL cholesterol	4085	142.85	0.03	0.06	0.01	0.41
	HDL cholesterol	4146	54.86	-0.04	0.01	-0.09	1E-8
	log2(Creatinine)	2748	-0.42	0.01	0.74	0.02	0.26
	Systolic blood pressure	4165	130.17	0.04	5E-3	0.07	4E-6
	Diastolic blood pressure	4165	75.86	0.05	3E-3	0.04	0.01
	log2(Waist / hip ratio)	4165	-0.28	0.05	3E-3	0.09	2E-8
BMI	4165	29.69	0.08	1E-6	0.09	2E-8	
Socio-behavioral	Education	4130	6.80	-0.02	0.14	-0.10	3E-10
	Income	4041	3.73	0.00	0.79	-0.06	1E-4
	log2(1+Exercise)	4142	2.53	-0.04	0.01	-0.07	2E-5
	Current smoker	4142	0.12	0.00	0.78	-0.01	0.66
	log2(1+Alcohol)	3687	1.10	-0.02	0.21	-0.07	3E-5

Figure 1. Marginal correlations with epigenetic age acceleration in the WHI. Correlations (bicor, biweight midcorrelation) between select variables and the two measures of epigenetic age acceleration are colored according to their magnitude with positive correlations in red, negative correlations in blue, and statistical significance (p-values) in green. Blood biomarkers were measured from fasting plasma collected at baseline. Food groups and nutrients are inclusive, including all types and all preparation methods, e.g. folic acid includes synthetic and natural, dairy includes cheese and all types of milk, etc. Variables are adjusted for ethnicity and dataset (BA23 or AS315).

Table 1. Characteristics of the WHI and InCHIANTI samples.

		WHI		InCHIANTI		
		Count	Percent	Count	Percent	
Ethnic	American Indian or Alaskan Native	56	1%			
	Asian or Pacific Islander	140	3%			
	Black or African-American	1277	28%			
	Hispanic/Latino	784	17%			
	White (not of Hispanic origin)	2196	49%			
	Other	37	1%			
WHI data set	BA23	2098	47%			
	AS315	2392	53%			
Sex	Male			178	44%	
	Female			229	56%	
Current smoker	Nonsmoker	4027	90%	367	90%	
	Smoker	439	10%	40	10%	
Education	< Primary	43	1%	80	20%	
	> Primary	154	3%	154	38%	
	> Lower secondary	293	7%	91	22%	
	> Upper secondary	2588	58%	62	15%	
	> Higher	1393	31%	20	5%	
Physical activity	Active	894	20%	329	81%	
	Inactive	3572	80%	78	19%	
		Mean	SD	Mean	SD	
Diet	Total energy, kcal	kcal/day	1641	777	2069	573
	Carbohydrate	% kcal	49.0	9.1	52.4	6.9
	Protein	% kcal	16.5	3.3	15.8	2.0
	Fat	% kcal	34.6	8.1	30.9	5.5
	Red meat	serv/day	0.8	0.7	1.1	0.5
	Poultry	serv/day	0.4	0.3	0.2	0.2
	Fish	serv/day	0.3	0.3	0.2	0.2
	Dairy	serv/day	1.6	1.3	2.8	1.8
	Whole grains	serv/day	1.2	0.9		
	Nuts	serv/day	0.2	0.3	0.0	0.1
	Fruits	serv/day	1.7	1.3	1.9	0.9
	Vegetables	serv/day	1.9	1.3	1.6	0.8
	Alcohol	g/day	3.6	9.6	12.7	14.9
	Measurements	IEAA	years	0.0	4.7	0.2
EEAA		years	0.0	6.0	-0.2	6.5
C-reactive protein		mg/L	5.2	6.6	3.9	7.4
Insulin		mg/dL	57.1	115.3		
Glucose		mg/dL	106.3	38.0	93.0	21.3
Triglycerides		mg/dL	146.4	85.6	122.7	81.5
Total cholesterol		mg/dL	228.4	42.7	207.2	36.6
LDL cholesterol		mg/dL	144.9	39.7	125.5	32.1
HDL cholesterol		mg/dL	54.0	14.3	57.6	15.7
Creatinine		mg/dL	0.8	0.2	0.9	0.4
Systolic blood pressure		mmHg	130.0	18.0	129.3	19.8
Diastolic blood pressure		mmHg	75.8	9.4	77.2	10.3
Waist / hip ratio		cm/cm	0.8	0.1	0.9	0.1
BMI	cm/m ²	29.7	6.0	27.0	4.3	

The cohort samples are listed for each column and variables of interest are listed for each row. The upper portion of the table correspond to categorical variables and are described using counts and percentages; the lower portion of the table displays continuous variables which are described using means and standard deviations (SD).

A	EEAA n	WHI										InCHIANTI 402	Meta-analysis			
		Caucasian 1807		African 1058		Hispanic 618		Asian 100					β	p	meta-t	meta-p
		β	p	β	p	β	p	β	p	β	p					
log2(1 + Fish)	-0.83	0.19	-1.05	0.12	-1.13	0.24	-0.55	0.78	-3.61	0.02	-2.92	3E-3				
log2(1 + Poultry)	-0.72	0.20	0.41	0.51	-0.15	0.83	2.06	0.29	-1.35	0.35	-0.73	0.46				
Current drinker	-0.66	0.02	-0.01	0.99	-1.58	9E-4	-2.13	0.07	-0.11	0.87	-3.23	1E-3				
Education	-0.19	0.01	-0.20	0.05	-0.23	0.03	-0.03	0.91	-0.13	0.06	-4.14	3E-5				
BMI	0.09	5E-4	0.10	3E-3	0.07	0.10	0.17	0.13	0.02	0.82	4.86	1E-6				
Physically active	-0.61	0.09	-0.11	0.82	-0.09	0.87	-2.69	0.08	-0.23	0.78	-1.70	0.09				
Current smoker	0.19	0.66	1.09	0.06	-0.80	0.23	-0.25	0.93	-2.34	0.02	0.07	0.94				

B	IEAA n	WHI										InCHIANTI 402	Meta-analysis			
		Caucasian 1807		African 1058		Hispanic 618		Asian 100					β	p	meta-t	meta-p
		β	p	β	p	β	p	β	p	β	p					
log2(1 + Fish)	0.61	0.24	-0.85	0.12	0.38	0.60	1.03	0.53	-2.51	0.03	-0.38	0.71				
log2(1 + Poultry)	-1.20	0.01	-0.15	0.76	-0.42	0.43	-2.28	0.17	-2.86	0.01	-3.30	1E-3				
Current drinker	-0.09	0.70	-0.11	0.75	-0.43	0.23	1.04	0.29	0.61	0.25	-0.36	0.72				
Education	0.01	0.84	-0.10	0.21	0.03	0.73	-0.34	0.16	-0.01	0.78	-0.68	0.49				
BMI	0.08	3E-5	0.04	0.17	0.02	0.60	0.35	5E-4	-0.03	0.56	4.14	4E-5				
Physically active	0.06	0.84	-0.40	0.28	-0.15	0.71	-2.22	0.09	-0.43	0.48	-1.06	0.29				
Current smoker	0.07	0.83	0.23	0.62	0.30	0.56	0.01	1.00	-1.25	0.11	0.12	0.91				

Figure 2. Meta-analysis of multivariable linear models of EEAA and IEAA in the WHI and InCHIANTI. EEAA (panel A) and IEAA (panel B) were regressed on potential confounding factors, fish and poultry intake, and current drinker status for the ethnic strata with sufficient sample sizes (n>100). Individual columns correspond to coefficient estimates (β) colored blue or red for negative and positive values respectively, and p-values (p) colored in green according to magnitude of significance, with the exception of the last two columns which denote Stouffer's method meta-t and meta-p values. Models are adjusted for originating dataset (WHI BA23 or AS315) and for sex (InCHIANTI).

Multivariate linear models were used to examine whether variations in cardiometabolic biomarkers and/or the number of symptoms for metabolic syndrome accounted for any of the associations between EEAA and lifestyle factors. The inclusion of biomarkers in an unstratified model shows that EEAA positively relates to CRP (log2, β=0.31, p=3x10⁻⁴, Figure 3A, model 3) and that this is accompanied by a concomitant diminishing in the effect size of BMI (67% decrease in coefficient magnitude, Figure 3A, model 2 vs. model 5), suggesting that higher CRP may partially explain the positive association between BMI and EEAA. When metabolic syndrome (MetS) was included in the model, results showed that higher EEAA is positively associated with the number of metabolic syndrome symptoms (β=0.29, p=0.002, Figure 3A, model 4). In the subset of participants with both biomarker and caro-

tenoid measurements, EEAA was negatively associated with mean carotenoid levels (β=-1.10, p=1x10⁻⁴) while appearing to diminish associations with biomarkers (Supplementary Figure 6A, model 5).

Additionally, we find that for the small subset of individuals for whom we have EEAA measurements at two time points (n=239, mean time interval = 2.7 years), increase in BMI (β=0.40, p=0.002) but not initial BMI (β=-0.01, p=0.81) is significantly associated with increased EEAA (higher follow-up EEAA after adjusting for the initial EEAA, dataset, and ethnicity).

Associations with IEAA

We conducted an analogous meta-analysis of ethnically-stratified linear models of IEAA and found that lower IEAA was significantly associated with poultry intake

A	EEAA n=2725	Model 1 Minimal		Model 2 Food		Model 3 Biomarkers		Model 4 MetS		Model 5 Full	
		β	p	β	p	β	p	β	p	β	p
		log2(1 + Fish)			-1.23	0.01					-1.13
log2(1 + Poultry)			-0.03	0.94					-0.05	0.90	
Current drinker			-0.63	0.01					-0.54	0.03	
Education	-0.27	7E-6	-0.22	2E-4	-0.24	5E-5	-0.25	2E-5	-0.20	8E-4	
BMI	0.09	1E-5	0.09	2E-5	0.02	0.46	0.06	0.01	0.03	0.26	
Physically active	-0.54	0.06	-0.44	0.13	-0.41	0.15	-0.49	0.08	-0.33	0.25	
Current smoker	0.30	0.38	0.33	0.32	0.23	0.50	0.26	0.43	0.28	0.41	
African American	-2.31	1E-17	-2.32	5E-17	-2.34	6E-16	-2.22	3E-16	-2.36	1E-15	
Hispanic	1.07	8E-4	0.98	2E-3	0.93	4E-3	1.11	5E-4	0.83	0.01	
log2(C-reactive protein)					0.31	3E-4			0.31	3E-4	
log2(Insulin)					0.19	0.25			0.19	0.24	
log2(Triglycerides)					0.25	0.24			0.34	0.14	
log2(Glucose)					0.34	0.32			0.41	0.27	
HDL Cholesterol					-0.01	0.33			-0.01	0.28	
Systolic blood pressure					0.01	0.13			0.01	0.08	
Diastolic blood pressure					0.00	0.95			0.00	0.75	
log2(Waist-to-hip ratio)					0.25	0.80			0.48	0.63	
Metabolic syndrome symptoms							0.29	2E-3	-0.16	0.30	

B	IEAA n=2725	Model 1 Minimal		Model 2 Food		Model 3 Biomarkers		Model 4 MetS		Model 5 Full	
		β	p	β	p	β	p	β	p	β	p
		log2(1 + Fish)			-0.04	0.92					0.01
log2(1 + Poultry)			-0.70	0.03					-0.71	0.03	
Current drinker			-0.07	0.72					0.00	0.99	
Education	-0.03	0.58	-0.02	0.72	-0.01	0.82	-0.01	0.75	-0.01	0.90	
BMI	0.06	2E-4	0.06	9E-5	0.02	0.19	0.03	0.07	0.03	0.20	
Physically active	-0.24	0.30	-0.23	0.31	-0.16	0.49	-0.19	0.39	-0.16	0.48	
Current smoker	0.26	0.33	0.26	0.34	0.26	0.33	0.23	0.38	0.25	0.35	
African American	-0.56	0.01	-0.50	0.02	-0.51	0.03	-0.48	0.03	-0.45	0.05	
Hispanic	-1.51	3E-9	-1.47	1E-8	-1.60	6E-10	-1.47	8E-9	-1.54	3E-9	
log2(C-reactive protein)					0.10	0.13			0.11	0.12	
log2(Insulin)					0.13	0.31			0.13	0.34	
log2(Triglycerides)					0.40	0.02			0.35	0.06	
log2(Glucose)					0.35	0.20			0.31	0.30	
HDL Cholesterol					0.00	0.89			0.00	0.72	
Systolic blood pressure					0.00	0.67			0.00	0.93	
Diastolic blood pressure					0.01	0.42			0.01	0.39	
log2(Waist-to-hip ratio)					-0.17	0.83			-0.24	0.77	
Metabolic syndrome symptoms							0.27	4E-4	0.07	0.58	

Figure 3. Multivariate linear models of EEAA and IEAA with and without biomarkers in the WHI. EEAA (panel A) and IEAA (panel B) were regressed on potential confounding factors, fish and poultry intake and current drinker status, and select biomarkers. Individual columns list the corresponding coefficient estimates (β) and p-values (p) for each fitting. Coefficients are colored according to sign (positive = red, negative = blue) and significance according to magnitude (green). Models 1 through 5 correspond to a minimal model, a model including dietary intake variables, a model including potential explanatory biomarkers, a model including number of metabolic syndrome symptoms and a complete model with all of the variables above, respectively. Models are adjusted for originating dataset (BA23 or AS315).

($t_{meta}=-3.30$, $p_{meta}=0.001$) and lower BMI ($t_{meta}=4.14$, $p_{meta}=4 \times 10^{-5}$), after adjusting for potential confounders (Figure 2B). In the subset of participants with measured carotenoids, IEAA was significantly associated with mean carotenoid levels ($t_{meta}=-2.47$, $p_{meta}=0.01$, Supplementary Figure 5B). When regressed on clinical biomarkers IEAA was significantly associated with triglycerides (\log_2 , $\beta=0.40$, $p=0.02$, Figure 3B, model 3). Their inclusion diminished the association between IEAA and BMI (60% decrease in coefficient magnitude, Figure 3B, model 2 vs. model 5). Number of metabolic syndrome symptoms was also significantly associated with IEAA ($\beta=0.27$, $p=4 \times 10^{-4}$, Figure 3B, model 4), and diminished the association between IEAA and BMI by 50%. In the subset of WHI participants with circulating carotenoid measurements, we find a trend toward association between IEAA and mean carotenoid levels ($\beta=-0.40$, $p=0.07$, Supplementary Figure 6B, model 5). Finally, in the participants with epigenetic profiling at two time points, increase in BMI ($\beta=0.22$, $p=0.03$) but not initial BMI ($\beta=-0.22$, $p=0.44$) is significantly associated with increased IEAA (higher follow-up IEAA after adjusting for the initial IEAA, dataset, and ethnicity).

Metformin and epigenetic age acceleration

Considering the significant association between our measures of epigenetic aging and markers of diabetes, we investigated the potential of metformin to modulate epigenetic aging. As shown in Supplementary Figure 7, women in the No Metformin group were the oldest, with mean ages around 65 years, compared to mean ages around 64 years for the Metformin Now group and 62.5 years for the Metformin Future group. Those in the No Metformin group also appeared to have a higher proportion of college graduates compared to the other two groups, whose members tended to have lower educational attainment. The two Metformin groups had much higher proportions of Non-Hispanic blacks, and somewhat higher proportions of Hispanics. As expected, women who were prescribed Metformin had the highest fasting insulin, fasting glucose, and HOMA-IR levels, with current Metformin users being the highest of all. Similarly, among Metformin users, fasting insulin, fasting glucose, and HOMA-IR levels were inversely associated with timing of Metformin initiation. This suggests that 1) women who had been on Metformin the longest at the time of blood draw were

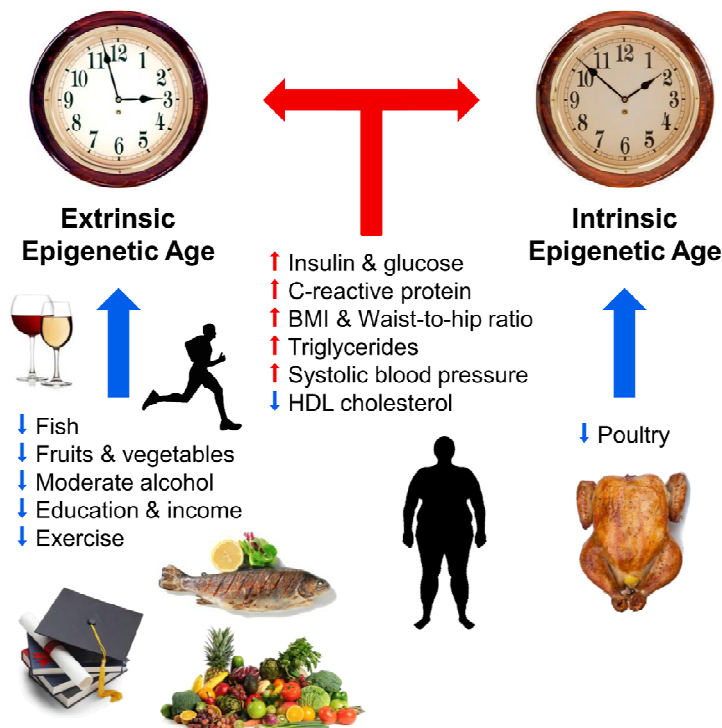


Figure 4. Pictorial summary of our main findings. The blue and red arrows depict anti-aging and pro-aging effects in blood respectively. The two clocks symbolize the extrinsic epigenetic clock (enhanced version of the Hannum estimate) and the intrinsic epigenetic clock (Horvath 2013) which are dependent and independent of blood cell counts, respectively.

the most insulin resistant—arguably because they were long-time diabetics—and 2) among women who would use Metformin in the future, those who would receive prescriptions sooner—who may already have pre-diabetes or diabetes at the time of blood draw—were more insulin resistant than women who were prescribed Metformin later.

As shown in Supplementary Figure 8, when examining the full sample, IEAA and EEAA were highest for women in the Metformin Now group, although differences between Metformin Now and Metformin Future were not significant. Further, the timing of Metformin prescription relative to the timing of blood draw was not associated with either IEAA or EEAA. Overall, these results held for IEAA when restricting the sample to only Non-Hispanic whites, Non-Hispanic blacks, or women at least 65 years of age. Results for EEAA held when restricting the sample to only Hispanics, women less than 65 years of age, or those with fasting glucose levels under 126 mg/dL. Nevertheless, there were no strata in which women in the Metformin Now group had significantly lower IEAA or EEAA than women in either the No Metformin or the Metformin Future group.

Given that Metformin may not be effective for all or that not all women prescribed Metformin may comply with their physician's recommendations, we compared future Metformin users to current Metformin users who had fasting glucose levels under 140 mg/dL (Supplementary Figure 9). Similar to Supplementary Figure 8, no significant differences were found for the two metformin groups, and in most cases those currently on Metformin had somewhat higher IEAA and EEAA than those who would be prescribed Metformin in the future.

Finally, when using women with repeat blood draws to compare the change in IEAA and EEAA for those who started Metformin between the two visits to those who didn't (Supplementary Table 1), we find that, while those who start Metformin appear to have lower IEAA ($\beta=-0.79$), the difference is not significant ($P=0.46$). Results for EEAA show very little difference between those who start Metformin between first and second blood draws, and those who don't ($\beta=-0.003$, $P=0.99$). Given that we only have 11 women who start Metformin, our null results, particularly for IEAA, may be due to a lack of power. Thus, in moving forward, this analysis may be worth revisiting in a randomized control trial or after acquiring larger sample sizes.

DISCUSSION

To our knowledge, this is the first study to examine associations between lifestyle factors and measures of

epigenetic age acceleration in blood. Our main findings are summarized graphically in Figure 4. Overall, our dietary results are consistent with some of the current Dietary Guidelines for Americans [60, 61], reflecting potential health benefits associated with higher intake of fish, poultry, and fruits and vegetables. The weak correlations between dietary factors and epigenetic aging rates probably reflect that a relatively large proportion of the variance in aging rates (around 40 percent) is explained by genetic factors [38, 59, 62]. We find that education, physical activity, low body mass index are associated with a slow extrinsic age acceleration both in univariate correlation tests (Figure 1) and in multivariate regression models (Figure 2A, 3A). However, consistent with our previous work, smoking status was not associated with epigenetic age acceleration [45], which highlights that not every poor lifestyle choice is associated with an increased epigenetic aging effect in blood tissue.

EEAA, inflammation, and metabolic functioning

The age-related changes in immune functioning and inflammation are believed to contribute to increased susceptibility of a wide range of diseases later in life, including diabetes, some cancers, cardiovascular, neurodegenerative, auto-immune, and infectious diseases [63, 64]. In our analysis, EEAA, a biomarker which explicitly incorporates aspects of immune system aging such as age-related changes in blood cell counts, was associated with cardiometabolic biomarkers, fish, fruit, vegetable, and alcohol intake.

Our finding that fish intake was negatively associated with EEAA is consistent with prospective studies suggesting that fish consumption is protective against various age-related diseases [65-67]. The benefits of fish intake may be mediated in part through the omega-3 fatty acids, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), which stimulate the synthesis of anti-inflammatory cytokines [68]. This is further supported by our finding that CRP—a well-known marker of inflammation—was the most significant explanatory biomarker of EEAA. This suggests that one reason higher fish consumption may lower EEAA is because it has beneficial anti-inflammatory or metabolic effects. The consensus between these associations also appears to converge on MetS as a potential mediating factor; this was further supported by our results showing that the number of MetS characteristics significantly relates to EEAA. Though CRP is not included in most MetS diagnostic criteria, the association between the two has been previously established [69].

We also find that alcohol consumption was negatively associated with EEAA even after adjusting for potential

confounders such as socioeconomic status; this is consistent with prospective studies which have identified light to moderate alcohol intake as a protective factor against all-cause and CHD-related mortality [70, 71] and is supported by a recent publication that also found an inverse association between epigenetic age and alcohol intake in Caucasian and African American individuals (n=656, n=180, respectively) [72]. In our study, we find that the potential benefits of alcohol consumption are observed using a threshold of more than one serving per month, though the effect size of this variable was also stable when adding weekly and daily intake levels (Supplementary Figure 4A-D). The association appears to be driven by wine consumption though there is also a trend towards association with beer (Supplementary Figures 4E-H). This is consistent with other studies have suggested that wine may have added benefits compared to light alcohol consumption [73]. This finding may also be related to the anti-inflammatory effects of light alcohol consumption, which are associated with decreased circulating levels of inflammatory markers such as IL-6 and CRP [74]. Alternatively, this may be the result of reverse causation, whereby individuals suffering from health issues abstain from alcohol consumption [75], though interventional studies support a causal protective effect of moderated alcohol intake on cardiovascular blood biomarkers [76].

Though EEAA only trended toward association with reported fruit and vegetable intake, we find significant associations with blood carotenoid levels, which are quantitative surrogates of fruit and vegetable intake; this is likely a reflection of the bias and inaccuracy of self-reported diet. This is in agreement with the wide range of literature supporting the protective effects of high fruit and vegetable intake against age-related diseases CHD [77, 78], stroke [79], type-2 diabetes [80], breast cancer [81], and all-cause mortality [82]. The association between fruit and vegetables with aging of the blood immune system may be partially mediated by anti-inflammatory [83, 84] and cardiometabolic effects, however it is interesting to note that the explanatory power of mean carotenoid levels remained even after including the other explanatory factors into the model, suggesting the possibility of independent anti-aging mechanisms (Supplementary Figure 6A, model 5).

Our results for EEAA also share similarities with previously reported findings showing that LTL relates to BMI [22], metabolic factors, vegetable consumption [85], and dietary intake of foods high in omega-3 fatty acids [86]. This agreement is likely a reflection of the shared immunological basis, which is supported by the weak negative correlation between EEAA and age-

adjusted LTL. In contrast, IEAA is not significantly associated with LTL, supporting the idea that these measures represent different aspects of aging.

Intrinsic epigenetic aging and metabolic health

Results showed that BMI has a positive association with IEAA (Figure 3B, model 2). The statistically significant but weak correlations between BMI and epigenetic age acceleration in blood ($r < 0.10$) are much smaller than those we recently reported for human liver ($r = 0.42$) [54], suggesting that associations between aging signatures and risk factors may vary in strength depending on the tissue, and may be stronger in organs/tissues most affected by the risk. Interestingly, IEAA was also associated with number of metabolic syndrome characteristics, suggesting a role in tracking metabolic aging processes (Figure 3B, model 4).

We did find that reported poultry intake was negatively associated with IEAA, even after adjusting for potential confounders and explanatory factors (Figure 3B, model 5). Given the relative inert behavior of IEAA, the mechanism by which poultry may affect aging is unclear.

Lack of an effect of metformin

Metformin, which is a first-line medication for the treatment of type 2 diabetes, has garnered substantial interest by aging researchers [87]. The fact that we did not detect an anti-aging effect of metformin in our study could be due to the following factors: a) it could reflect the limitations/biases of an observational study, b) it could reflect the limited sample size, or c) that metformin does not affect epigenetic aging rates of blood tissue. Future randomized controlled trials should revisit the question whether metformin affects epigenetic aging rates in blood or other tissues.

Generalization to the InCHIANTI

Our results from the InCHIANTI show some validation of our findings from the WHI: fish intake was related to EEAA, and poultry was related to IEAA. Associations with available biomarkers of cardiometabolic health, however, were not found to be validated, and in a few cases were reversed in directions, within the InCHIANTI (data not shown).

The discrepancies between the WHI and InCHIANTI cohorts may be due to numerous differences in the study population (cultural, demographic, genetic, health status, Table 1) and data collection methodology (dietary assessment). Despite being younger, on average, US participants from the WHI had higher body

mass indexes (BMI), and worse metabolic health than their Italian counterparts—as indicated by their greater prevalence of metabolic syndrome (23% in the WHI versus 7.6% in the InCHIANTI).

The InCHIANTI study is also arguably underpowered ($n=402$) when it comes to detecting the weak associations with epigenetic age acceleration. According to sample size calculations (*PASS* software), we find that $n=1820$ samples are needed to provide 80% power to detect a correlation of $r=0.08$ at a two-sided significance level of $\alpha=0.01$. Similarly, $n=1163$ samples are needed to detect a correlation of $r=0.10$.

Limitations

While our study of the WHI benefits from having a relatively large sample size, associations with epigenetic aging may not be detectable in smaller studies given the weak effect sizes observed here. This situation is exacerbated by self-reported lifestyle habits which are notorious for bias and inaccuracy, limiting their ability to represent true lifestyle habits and potentially producing false negative results. Further, as evidenced by our results from the ethnic strata and the InCHIANTI, studies conducted in different ethnic populations may not be entirely consistent due to fundamental differences in age, diet, culture, and demographics. There were several potential limitations to this study, which include the assumption of non-confounding from unmeasured variables such as existing patient co-morbidities and the assumption of accuracy and long-term consistency of reported dietary habits. This is the first longitudinal study to show that an increase in BMI is associated with an increase in epigenetic age acceleration but larger longitudinal studies will be needed to dissect causal relationships between epigenetic aging rates and dietary measures, education, exercise, and lifestyle factors.

Conclusions about epigenetic age acceleration

Our large sample size ($n>4500$) provides sufficient statistical power for one of our main conclusions: diet has only a weak effect on epigenetic aging rates in blood. These findings will be valuable for researchers who plan to use epigenetic biomarkers in dietary intervention studies. The wide range of associations found with EEAA suggest that immune system aging may be closely linked to conventional notions of metabolic health and may be sensitive to variations in environment and lifestyle. In contrast, IEAA has few associations, which is consistent with the hypothesis that cell-intrinsic aging remains relatively stable, more likely being determined by an intrinsic aging or developmental process under genetic control. Further,

using longitudinal data in the WHI, we found that change in both EEAA and IEAA are significantly associated with change in BMI, suggesting that both modes of epigenetic aging may respond to changes in lifestyle, at least with respect to change in obesity. Overall, our results are consistent with previous literature supporting the protective effects of fish, poultry and alcohol consumption, exercise, education, as well as the risk of obesity and dyslipidemia.

METHODS

Estimation of DNA methylation age

DNAm age (also referred to as epigenetic age) was calculated from human samples profiled with the Illumina Infinium 450K platform, described in detail in [38]. Briefly, the epigenetic clock is defined as a prediction method of age based on the DNAm levels of 353 CpGs. Predicted age, referred to as DNAm age, correlates with chronological age in sorted cell types (CD4+ T cells, monocytes, B cells, glial cells, neurons), tissues, and organs, including: whole blood, brain, breast, kidney, liver, lung, saliva [38].

We also applied the Hannum measure of DNAm age based on 71 CpGs which was developed using DNA methylation data from blood [37].

Despite high correlations, DNAm age estimates can deviate substantially from chronological age at the individual level, and adjusting for age we can arrive at measures of epigenetic age acceleration as described in the following.

Estimation of Intrinsic and Extrinsic Epigenetic Age Acceleration (IEAA, EEAA)

In this article, we consider two measures of epigenetic age acceleration. These measures, referred to as intrinsic and extrinsic age acceleration only apply to blood. IEAA is derived from the Horvath measure of DNAm age based on 353 CpGs [38], and is defined as the residual resulting from regressing Horvath DNAm age on chronological age and estimates of plasmablasts, naive and exhausted CD8+ T cells, CD4+ T cells, natural killer cells, monocytes, and granulocytes. Thus, IEAA is independent of chronological age and most of the variation in blood cell composition. IEAA is meant to capture cell-intrinsic properties of the aging process that exhibits preservation across various cell types and organs.

EEAA can be interpreted as an enhanced version of the Hannum measure of DNAm age based on 71 CpGs [37]. EEAA up-weights the contributions of age related

blood cell counts [43]. Specifically, EEAA is defined using the following three steps. First, we calculated the epigenetic age measure from Hannum et al, which already correlated with certain blood cell types [40]. Second, we increased the contribution of immune blood cell types to the age estimate by forming a weighted average of Hannum's estimate with 3 cell types that are known to change with age: naïve (CD45RA+CCR7+) cytotoxic T cells, exhausted (CD28-CD45RA-) cytotoxic T cells, and plasmablasts using the Klemra Doubal approach [88]. The weights used in the weighted average are determined by the correlation between the respective variable and chronological age. The weights were chosen on the basis of the WHI data and the same (static) weights were used for all data sets. Finally, EEAA was defined as the residual variation resulting from a univariate model regressing the resulting age estimate on chronological age. Thus, EEAA tracks both age related changes in blood cell composition and intrinsic epigenetic changes.

In a recent large scale meta-analysis involving over 13 thousand subjects from 13 cohorts, we have shown that both IEAA and EEAA are predictive of mortality, independent of chronological age, even after adjusting for additional risk factors, and within the racial/ethnic groups that we examined (Caucasians, Hispanics, African Americans) [43].

IEAA and EEAA can be obtained from the online DNAm age calculator (<http://labs.genetics.ucla.edu/horvath/dnamage/>), where they are denoted as *AAHOAdjCellCounts* and *BioAge4HASStaticAdjAge*, respectively.

Dietary assessment in the Women's Health Initiative (WHI)

Participants were selected from the WHI, a national study that began in 1993 and enrolled postmenopausal women between the ages of 50-79 years into either randomized clinical trials (RCTs) or into an observational study [89] (more details available in Supplementary Information). Participants completed self-administered questionnaires at baseline which provided personal information on a wide range of topics, including sociodemographic information (age, education, race, income), and current health behaviors (recreational physical activity, tobacco and alcohol exposure, and diet). Participants also visited clinics at baseline where certified Clinical Center staff collected blood specimens and performed anthropometric measurements including weight, height, hip and waist circumferences, and systolic and diastolic blood pressures; body mass index and waist to hip ratio were calculated from these measurements (Table 1).

Dietary intake levels were assessed at baseline using the WHI Food Frequency Questionnaire [90]. Briefly, participants were asked to report on dietary habits in the past three months, including intake, frequency, and portion sizes of foods or food groups, along with questions concerning topics such as food preparation practices and types of added fats. Nutrient intake levels were then estimated from these responses. For current drinker, we use the threshold of more than one serving equivalent (14g) within the last 28 days.

Estimation of blood cell counts based on DNA methylation levels

We estimate blood cell counts using two different software tools. First, Houseman's estimation method [91], which is based on DNA methylation signatures from purified leukocyte samples, was used to estimate the proportions of CD8+ T cells, CD4+ T, natural killer, B cells, and granulocytes (also known as polymorphonuclear leukocytes). Second, the advanced analysis option of the epigenetic clock software [38, 52] was used to estimate the percentage of exhausted CD8+ T cells (defined as CD28-CD45RA-) and the number (count) of naïve CD8+ T cells (defined as CD45RA+CCR7+). We and others have shown that the estimated blood cell counts have moderately high correlations with corresponding flow cytometric measures [91, 92]. For example, flow cytometric measurements from the MACS study correlate strongly with DNA methylation based estimates: $r=0.63$ for CD8+T cells, $r=0.77$ for CD4+ T cells, $r=0.67$ for B cell, $r=0.68$ for naïve CD8+ T cell, $r=0.86$ for naïve CD4+ T, and $r=0.49$ for exhausted CD8+ T cells [92].

Blood biomarkers and DNA methylation in the WHI

Two separate subsamples were aggregated for our study within the WHI (BA23 and AS315). Both had baseline blood specimens collected after an overnight fast in EDTA tubes and stored at -70C. These samples were processed at the WHI core laboratory and select nutrient and cardiovascular biomarkers were measured including lycopene, alpha- & beta-carotene, alpha- & gamma-tocopherol, C-reactive protein, triglycerides, total, LDL, and HDL cholesterol.

For the first subsample (BA23) consisting of 2098 samples, DNA methylation levels were measured using the Illumina Infinium HumanMethylation450 BeadChip at the HudsonAlpha Institute of Biotechnology. This platform uses bisulfite conversion to quantify methylation levels at 485,577 specific CpG sites genome-wide. Samples were prepared according to the standard Illumina protocol, and β methylation values were calculated from the intensity ratio between

methylated and total (methylated and unmethylated) probe fluorescence intensities. Methylation data was processed as described in [38]. In order to test the quality of these array measurements, we perform correlation measures with duplicates within this dataset and with a "gold" standard which is an average of many samples previously collected. Correlation between duplicates and with the gold standard were high ($r > 0.9$), indicative of high quality measurements.

The second WHI data set is described in the following.

WHI-EMPC description

The Women's Health Initiative – Epigenetic Mechanisms of PM-Mediated CVD (WHI-EMPC, AS315) is an ancillary study of epigenetic mechanisms underlying associations between ambient particulate matter (PM) air pollution and cardiovascular disease (CVD) in the Women's Health Initiative clinical trials (CT) cohort. The WHI-EMPC study population is a stratified, random sample of 2,200 WHI CT participants who were examined between 1993 and 2001; had available buffy coat, core analytes, electrocardiograms, and ambient concentrations of PM; but were not taking anti-arrhythmic medications at the time. As such, WHI-EMPC is representative of the larger, multiethnic WHI CT population from which it was sampled: $n = 68,132$ participants aged 50-79 years who were randomized to hormone therapy, calcium/vitamin D supplementation, and/or dietary modification in 40 U.S. clinical centers at the baseline exam (1993-1998) and re-examined in the fasting state one, three, six, and nine years later [93].

Illumina Infinium HumanMethylation450 BeadChip data from the Northwestern University Genomics Core Facility for WHI-EMPC participants sampled in stages 1a (800 participants), 1b (1200 participants), and 2 (200 participants x 2 samples each) was quality controlled and batch adjusted. Batch adjustment involved applying empirical Bayes methods of adjusting for stage and plate as implemented in ComBat [94].

Dietary assessment in the Invecchiare nel Chianti (InCHIANTI)

The InCHIANTI Study is a population-based prospective cohort study of residents ages 30 or older from two areas in the Chianti region of Tuscany, Italy. Data on demographic and lifestyle factors such as smoking, years of education, BMI, and physical activity were collected during the baseline interview. Physical activity in the previous year was categorized as sedentary or active. Smoking was categorized into current smoker versus former or non-smokers (Table 1).

In the InCHIANTI study, dietary intake for the past year was assessed using a 236 item food frequency questionnaire (FFQ) for the European Prospective Investigation on Cancer and nutrition (EPIC) study, previously validated in the InCHIANTI population [95]. The FFQ was administered by a trained interviewer and collected information on how frequently (weekly, monthly, yearly) each specific food was generally consumed. Participants were asked to specify the size of the portion usually consumed, in comparison to a range of portion that are shown in colored photographs. Nutrient data for specific foods were obtained from the Food Composition Database for Epidemiological Studies in Italy [96]. Dietary information was judged as unreliable and excluded from further analysis if reported energy intakes were < 600 kcal/day or $> 4,000$ kcal/day and $> 4,200$ kcal/day in women and men, respectively.

Blood biomarkers and DNA methylation in the InCHIANTI

Sampling and data collection procedures have been described elsewhere [97]. Briefly, participants were enrolled between 1998 and 2000 and were examined at three-year intervals. Serum samples obtained from blood collected in evacuated tubes without anti-coagulant were centrifuged at 2000g for 10 min, and stored at -80 °C for measurement of glucose, total, LDL, and HDL, cholesterol, triglycerides, CRP, and creatinine. DNA methylation was assayed using the Illumina Infinium HumanMethylation450 platform for $n = 407$ participants with sufficient DNA at both baseline (years 1998-2000) and year 9 follow-up visits (2007-2009).

Assessment of metabolic syndrome

Metabolic syndrome status was assessed using the ATP III NCEP 2004 criteria defined by the presence of 3 or more of the following characteristics: waist circumference > 88 cm (if male, > 102 cm), systolic blood pressure > 130 mmHg or diastolic blood pressure > 85 mmHg, fasting plasma glucose > 100 mg/dL, HDL cholesterol < 50 mg/dL (if male, < 40 mg/dL), and triglycerides > 150 mg/dL. In regression models, we use total number of metabolic syndrome characteristics as an ordinal variable, ranging from 0 to 5.

Statistical analyses

Dietary analysis

Biweight midcorrelation, an outlier-robust correlation measure, was used to assess marginal linear relationships between epigenetic aging measures and dietary, cardiometabolic, and socioeconomic factors. To

adjust for possible socioeconomic and lifestyle confounders, we fit ethnically-stratified multivariable linear models adjusting for education, exercise, BMI, and current drinker and smoker status. We used Stouffer's method to infer the meta-analytic significance of each variable over the different ethnic strata using the square-root of the sample size as the Z-score weighting factor. Specifically for the WHI, the age acceleration measures were adjusted for differences in originating dataset and within the InCHIANTI the measures were adjusted for sex. Models including regression on biomarkers, and number of metabolic syndrome symptoms were not stratified by ethnicity due to lack of coverage for biomarker profiling. Models were designed based on common prior knowledge and in cases where there was co-linearity between confounding variables, choice for adjustment was selected based on variable commonality in order to improve comparability with other studies, e.g. BMI was chosen over WHR because BMI is more commonly measured and reported. Variables with skewness >1 were log transformed (possibly adding +1 to avoid forming the logarithm of zero). Mean carotenoids was computed as the mean across standardized measures of lycopene, log₂(alpha-carotene), log₂(beta-carotene), log₂(lutein + zeaxanthin), and log₂(beta-cryptoxanthin). Repeat measurements on the same individuals were omitted from the analysis.

Metformin analysis

Prescription medication use was collected at baseline for women in both the clinical trial and the observational study. Follow-up recording of new medication was done at years 1, 3, 6, and 9 for women enrolled in clinical trial and at year 3 for those enrolled in the observational study. At each of these waves, participants were asked to bring in their prescription medication bottles and interviewers recorded the National Drug Code (NDC), the product name, and generic name for each drug. In our analysis, those who took prescription drugs that fell under the Therapeutic Class Codes (TCCODE), 272500 or 279970 were coded as taking Metformin. TCCODE 272500 covered the following drugs: Metformin HCL, Metformin ER, Fortamet, Glucophage XR, Glucophage, and Riomet; all of which were Metformin HCL tablets. TCCODE 279970 covered the following drugs: Glucovance, Glyburide/Metformin HCL, and Metaglip; the first two of which were Glyburide and Metformin HCL combination tablets, whereas the later is a Glipizide Metformin HCL combination tablet. In our study, approximately 12% (n=489) of women were reported as using Metformin at some time during the study. Women were also asked the age at which they started using prescribed medication. On average, women in our study began metformin use at age 65, with the youngest and

oldest ages of initiation reported as 41 and 85, respectively.

We utilized the reported age of Metformin initiation to estimate whether women were using metformin at the time of blood draw (when DNAm age was measured), and if so, how long they had been on the drug. For those women who started using Metformin after the date of blood draw, we were also able to calculate the time between blood draw and starting on Metformin. Based on this, we classified women in to three categories—“No Metformin” (those who never used Metformin), “Metformin Now” (those on Metformin at the time of blood draw), and “Metformin Future” (those who began Metformin after blood draw). In our sample, 4,073 women were in the No Metformin group, 139 were in the Metformin Now group, and 350 were in the Metformin Future group.

Kruskal Wallis tests were used to compare the ages at blood draw, education levels, race/ethnicity, fasting glucose levels, fasting insulin levels, and degree of insulin resistance (measured using the Homeostatic Model Assessment of Insulin Resistance--HOMA_IR). Biweight midcorrelations were used to examine whether these characteristics differed as a function of the timing of Metformin initiation among those on Metformin at some point during the study. Next Kruskal Wallis tests and biweight midcorrelations were used to examine whether Metformin group classification and/or the timing of Metformin use was associated with IEAA or EEAA. For these models, both IEAA and EEAA were adjusted for age at blood draw, education, and race/ethnicity. These models were run using the full sample, and stratifying by age at blood draw (<65 years vs. 65+ years), race/ethnicity, and fasting glucose levels (<126 mg/dL vs. 126+ mg/dL). Next, a two group comparison was made for the full sample and stratifying by race/ethnicity between a) 47 women in the Metformin Now group for whom Metformin was potentially effective (fasting glucose < 140 mg/dL), and b) the 350 women in the Metformin Future group. Finally, there were 308 women who 1) had IEAA and EEAA measured during at least two separate WHI visits and 2) were not on Metformin at the time of their first blood draw. Of these women, 11 started Metformin in between their two blood draws. Based on this, we used linear models to compare the change in IEAA and EEAA between women who started Metformin between their first and second IEAA and EEAA measure and those who did not. Models were adjusted for age at first blood draw, age at second blood draw, race/ethnicity, education, glucose levels at first blood draw, and either IEAA or EEAA at first blood draw.

AUTHOR CONTRIBUTIONS

Austin Quach, Morgan E. Levine, Toshiko Tanaka, Brian H. Chen, and Steve Horvath carried out the analysis. Devin Absher, Themistocles L. Assimes, Phil Tsao, Steve Horvath, Beate Ritz, Luigi Ferrucci, Stefania Bandinelli, Eric A. Whitsel, Lifang Hou, Andrea Baccarelli, Yun Li, James Stewart contributed data. Marian L. Neuhouser, Jeannette M. Beasley, Linda Snetselaar, Robert B. Wallace, Ake Lu helped with the interpretation of the findings and drafting of the article. Steve Horvath designed the study.

ACKNOWLEDGEMENTS

We would like to acknowledge The WHI Investigators listed below:

Program Office: (National Heart, Lung, and Blood Institute, Bethesda, Maryland) Jacques Rossouw, Shari Ludlam, Dale Burwen, Joan McGowan, Leslie Ford, and Nancy Geller.

Clinical Coordinating Center: (Fred Hutchinson Cancer Research Center, Seattle, WA) Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg.

Investigators and Academic Centers: (Brigham and Women's Hospital, Harvard Medical School, Boston, MA) JoAnn E. Manson; (MedStar Health Research Institute/Howard University, Washington, DC) Barbara V. Howard; (Stanford Prevention Research Center, Stanford, CA).

Marcia L. Stefanick; (The Ohio State University, Columbus, OH) Rebecca Jackson; (University of Arizona, Tucson/Phoenix, AZ) Cynthia A. Thomson; (University at Buffalo, Buffalo, NY)

Jean Wactawski-Wende; (University of Florida, Gainesville/Jacksonville, FL) Marian Limacher; (University of Iowa, Iowa City/Davenport, IA) Robert Wallace; (University of Pittsburgh, Pittsburgh, PA) Lewis Kuller; (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker.

Women's Health Initiative Memory Study: (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker.

CONFLICTS OF INTEREST

The Regents of the University of California is the sole owner of a provisional patent application directed at the invention of measures of epigenetic age acceleration for which SH is a named inventor. The other authors declare no conflict of interest.

FUNDING

This WHI study was supported by National Institutes of Health NIH/NHLBI 60442456 BAA23 (Assimes,

Absher, Horvath) and the National Institute of Environmental Health Sciences R01ES020836 WHI-EMPC (Whitsel, Baccarelli, Hou), R01ES021733 (Baccarelli), and R01ES025225 (Baccarelli). The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf.

Additional support came from NIH/NIA 5R01AG042511-02 (Horvath), NIH/NIA U34AG051425-01 (Horvath), NIH/NINDS T32NS048004 (Levine), and the Burroughs Wellcome Fund Inter-school Training Program in Chronic Diseases (BWF-CHIP, Quach). The funding bodies played no role in the design, the collection, analysis, or interpretation of the data.

The InCHIANTI study baseline (1998-2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336).

REFERENCES

1. Kaczmarczyk MM, Miller MJ, Freund GG. The health benefits of dietary fiber: beyond the usual suspects of type 2 diabetes mellitus, cardiovascular disease and colon cancer. *Metabolism*. 2012; 61:1058–66. doi: 10.1016/j.metabol.2012.01.017
2. Kris-Etherton PM, Harris WS, Appel LJ, and American Heart Association. Nutrition Committee. Fish consumption, fish oil, omega-3 fatty acids, and cardiovascular disease. *Circulation*. 2002; 106:2747–57. doi: 10.1161/01.CIR.0000038493.65177.94
3. van't Veer P, Jansen MC, Klerk M, Kok FJ. Fruits and vegetables in the prevention of cancer and cardiovascular disease. *Public Health Nutr*. 2000; 3:103–07. doi: 10.1017/S1368980000000136
4. Giugliano D, Ceriello A, Esposito K. The effects of diet on inflammation: emphasis on the metabolic syndrome. *J Am Coll Cardiol*. 2006; 48:677–85. doi: 10.1016/j.jacc.2006.03.052
5. Mozaffarian D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity. 2016; 133:

- 187–225. doi: 10.1161/CIRCULATIONAHA.115.018585
6. Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, Fruchart JC, James WP, Loria CM, Smith SC Jr, and International Diabetes Federation Task Force on Epidemiology and Prevention, and National Heart, Lung, and Blood Institute, and American Heart Association, and World Heart Federation, and International Atherosclerosis Society, and International Association for the Study of Obesity. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation*. 2009; 120:1640–45. doi: 10.1161/CIRCULATIONAHA.109.192644
 7. Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, Gordon DJ, Krauss RM, Savage PJ, Smith SC Jr, Spertus JA, Costa F, and American Heart Association, and National Heart, Lung, and Blood Institute. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. *Circulation*. 2005; 112:2735–52. doi: 10.1161/CIRCULATIONAHA.105.169404
 8. Wilson PW, D'Agostino RB, Parise H, Sullivan L, Meigs JB. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. *Circulation*. 2005; 112:3066–72. doi: 10.1161/CIRCULATIONAHA.105.539528
 9. Thun MJ, Peto R, Lopez AD, Monaco JH, Henley SJ, Heath CW Jr, Doll R. Alcohol consumption and mortality among middle-aged and elderly U.S. adults. *N Engl J Med*. 1997; 337:1705–14. doi: 10.1056/NEJM199712113372401
 10. Rimm EB, Williams P, Fosher K, Criqui M, Stampfer MJ. Moderate alcohol intake and lower risk of coronary heart disease: meta-analysis of effects on lipids and haemostatic factors. *BMJ*. 1999; 319:1523–28. doi: 10.1136/bmj.319.7224.1523
 11. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. *CMAJ*. 2006; 174:801–09. doi: 10.1503/cmaj.051351
 12. González MA, Rodríguez Artalejo F, Calero JR. Relationship between socioeconomic status and ischaemic heart disease in cohort and case-control studies: 1960-1993. *Int J Epidemiol*. 1998; 27:350–58. doi: 10.1093/ije/27.3.350
 13. Madsen M, Andersen AM, Christensen K, Andersen PK, Osler M. Does educational status impact adult mortality in Denmark? A twin approach. *Am J Epidemiol*. 2010; 172:225–34. doi: 10.1093/aje/kwq072
 14. Conti G, Heckman J, Urzua S. THE EDUCATION-HEALTH GRADIENT. *Am Econ Rev*. 2010; 100:234–38. doi: 10.1257/aer.100.2.234
 15. Currie J. Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. National Bureau of Economic Research Working Paper Series. 2008; No. 13987.
 16. Cutler DM, Lleras-Muney A. Understanding differences in health behaviors by education. *J Health Econ*. 2010; 29:1–28. doi: 10.1016/j.jhealeco.2009.10.003
 17. Mirabello L, Huang WY, Wong JY, Chatterjee N, Reding D, Crawford ED, De Vivo I, Hayes RB, Savage SA. The association between leukocyte telomere length and cigarette smoking, dietary and physical variables, and risk of prostate cancer. *Aging Cell*. 2009; 8:405–13. doi: 10.1111/j.1474-9726.2009.00485.x
 18. Boccardi V, Esposito A, Rizzo MR, Marfella R, Barbieri M, Paolisso G. Mediterranean diet, telomere maintenance and health status among elderly. *PLoS One*. 2013; 8:e62781. doi: 10.1371/journal.pone.0062781
 19. García-Calzón S, Gea A, Razquin C, Corella D, Lamuela-Raventós RM, Martínez JA, Martínez-González MA, Zalba G, Martí A. Longitudinal association of telomere length and obesity indices in an intervention study with a Mediterranean diet: the PREDIMED-NAVARRA trial. *Int J Obes*. 2014; 38:177–82. doi: 10.1038/ijo.2013.68
 20. Brouillette SW, Moore JS, McMahon AD, Thompson JR, Ford I, Shepherd J, Packard CJ, Samani NJ, and West of Scotland Coronary Prevention Study Group. Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. *Lancet*. 2007; 369:107–14. doi: 10.1016/S0140-6736(07)60071-3
 21. Kiecolt-Glaser JK, Epel ES, Belury MA, Andridge R, Lin J, Glaser R, Malarkey WB, Hwang BS, Blackburn E. Omega-3 fatty acids, oxidative stress, and leukocyte telomere length: A randomized controlled trial. *Brain Behav Immun*. 2013; 28:16–24. doi: 10.1016/j.bbi.2012.09.004
 22. Cassidy A, De Vivo I, Liu Y, Han J, Prescott J, Hunter DJ, Rimm EB. Associations between diet, lifestyle factors, and telomere length in women. *Am J Clin Nutr*. 2010; 91:1273–80. doi: 10.3945/ajcn.2009.28947
 23. Latifovic L, Peacock SD, Massey TE, King WD. The Influence of Alcohol Consumption, Cigarette Smoking,

- and Physical Activity on Leukocyte Telomere Length. *Cancer Epidemiol Biomarkers Prev.* 2016; 25:374–80. doi: 10.1158/1055-9965.EPI-14-1364
24. Cherkas LF, Hunkin JL, Kato BS, Richards JB, Gardner JP, Surdulescu GL, Kimura M, Lu X, Spector TD, Aviv A. The association between physical activity in leisure time and leukocyte telomere length. *Arch Intern Med.* 2008; 168:154–58. doi: 10.1001/archinternmed.2007.39
 25. Adler N, Pantell MS, O'Donovan A, Blackburn E, Cawthon R, Koster A, Opresko P, Newman A, Harris TB, Epel E. Educational attainment and late life telomere length in the Health, Aging and Body Composition Study. *Brain Behav Immun.* 2013; 27:15–21. doi: 10.1016/j.bbi.2012.08.014
 26. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh RF, Wiencke JK, Kelsey KT. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.* 2009; 5:e1000602. doi: 10.1371/journal.pgen.1000602
 27. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P, Baccarelli A. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev.* 2009; 130:234–39. doi: 10.1016/j.mad.2008.12.003
 28. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.* 2010; 20:434–39. doi: 10.1101/gr.103101.109
 29. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, Savage DA, Mueller-Holzner E, Marth C, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* 2010; 20:440–46. doi: 10.1101/gr.103606.109
 30. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* 2012; 13:R97. doi: 10.1186/gb-2012-13-10-r97
 31. Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, Wininger M, Colantuoni C, Weinberger DR, Kleinman JE, Lipska BK. DNA methylation signatures in development and aging of the human prefrontal cortex. *Am J Hum Genet.* 2012; 90:260–72. doi: 10.1016/j.ajhg.2011.12.020
 32. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, Warren ST. Age-associated DNA methylation in pediatric populations. *Genome Res.* 2012; 22:623–32. doi: 10.1101/gr.125187.111
 33. Johansson A, Enroth S, Gyllensten U. Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *PLoS One.* 2013; 8:e67378. doi: 10.1371/journal.pone.0067378
 34. Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD, Myers RM, Absher D. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biol.* 2013; 14:R102. doi: 10.1186/gb-2013-14-9-r102
 35. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E. Epigenetic predictor of age. *PLoS One.* 2011; 6:e14821. doi: 10.1371/journal.pone.0014821
 36. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S, Rezzi S, Castellani G, Capri M, et al. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell.* 2012; 11:1132–34. doi: 10.1111/accel.12005
 37. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013; 49:359–67. doi: 10.1016/j.molcel.2012.10.016
 38. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013; 14:R115. doi: 10.1186/gb-2013-14-10-r115
 39. Lin Q, Weidner CI, Costa IG, Marioni RE, Ferreira MR, Deary IJ, Wagner W. DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. *Aging (Albany NY).* 2016; 8:394–401. doi: 10.18632/aging.100908
 40. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* 2015; 16:25. doi: 10.1186/s13059-015-0584-6
 41. Christiansen L, Lenart A, Tan Q, Vaupel JW, Aviv A, McGue M, Christensen K. DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell.* 2016; 15:149–54. doi: 10.1111/accel.12421
 42. Perna L, Zhang Y, Mons U, Hollecsek B, Saum K-U, Brenner H. Epigenetic age acceleration predicts

- cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics*. 2016; 8:64. doi: 10.1186/s13148-016-0228-z
43. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, Roetker NS, Just AC, Demerath EW, Guan W, Bressler J, Fornage M, Studenski S, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)*. 2016; 8:1844–65. doi: 10.18632/aging.101020.
44. Breitling LP, Saum KU, Perna L, Schöttker B, Holleczeck B, Brenner H. Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clin Epigenetics*. 2016; 8:21. doi: 10.1186/s13148-016-0186-5
45. Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)*. 2015; 7:690–700. doi: 10.18632/aging.100809
46. Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, Gibson J, Redmond P, Cox SR, Pattie A, Corley J, Taylor A, Murphy L, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol*. 2015; 44:1388–96. doi: 10.1093/ije/dyu277
47. Horvath S, Pirazzini C, Bacalini MG, Gentilini D, Di Blasio AM, Delledonne M, Mari D, Arosio B, Monti D, Passarino G, De Rango F, D'Aquila P, Giuliani C, et al. Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. *Aging (Albany NY)*. 2015; 7:1159–70. doi: 10.18632/aging.100861
48. Levine ME, Lu AT, Bennett DA, Horvath S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany NY)*. 2015; 7:1198–211. doi: 10.18632/aging.100864
49. Horvath S, Mah V, Lu AT, Woo JS, Choi OW, Jasinska AJ, Riancho JA, Tung S, Coles NS, Braun J, Vinters HV, Coles LS. The cerebellum ages slowly according to the epigenetic clock. *Aging (Albany NY)*. 2015; 7:294–306. doi: 10.18632/aging.100742
50. Walker RF, Liu JS, Peters BA, Ritz BR, Wu T, Ophoff RA, Horvath S. Epigenetic age analysis of children who seem to evade aging. *Aging (Albany NY)*. 2015; 7:334–39. doi: 10.18632/aging.100744
51. Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D, Di Blasio AM, Giuliani C, Tung S, Vinters HV, Franceschi C. Accelerated epigenetic aging in Down syndrome. *Aging Cell*. 2015; 14:491–95. doi: 10.1111/acel.12325
52. Horvath S, Levine AJ. HIV-1 infection accelerates age according to the epigenetic clock. *J Infect Dis*. 2015; 212:1563–73. doi: 10.1093/infdis/jiv277
53. Horvath S, Langfelder P, Kwak S, Aaronson J, Rosinski J, Vogt TF, Eszes M, Faull RL, Curtis MA, Waldvogel HJ, Choi OW, Tung S, Vinters HV, et al. Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging (Albany NY)*. 2016; 8:1485–512. doi: 10.18632/aging.101005
54. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schönfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD, Deloukas P, Siebert R, Sipos B, et al. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci USA*. 2014; 111:15538–43. doi: 10.1073/pnas.1412759111
55. Zannas AS, Arloth J, Carrillo-Roa T, Iurato S, Röh S, Ressler KJ, Nemeroff CB, Smith AK, Bradley B, Heim C, Menke A, Lange JF, Brückl T, et al. Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol*. 2015; 16:266. doi: 10.1186/s13059-015-0828-5
56. Levine ME, Lu AT, Chen BH, Hernandez DG, Singleton AB, Ferrucci L, Bandinelli S, Salfati E, Manson JE, Quach A, Kusters CD, Kuh D, Wong A, et al. Menopause accelerates biological aging. *Proc Natl Acad Sci USA*. 2016; 113:9327–32. doi: 10.1073/pnas.1604558113
57. Vidal L, Lopez-Golan Y, Rego-Perez I, Horvath S, Blanco FJ, Riancho JA, Gomez-Reino JJ, Gonzalez A. Specific increase of methylation age in osteoarthritis cartilage. *Osteoarthritis Cartilage*. 2016; 24:S63. doi: 10.1016/j.joca.2016.01.140
58. Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)*. 2015; 7:1130–42. doi: 10.18632/aging.100859
59. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, Ritz BR, Chen B, Lu AT, Rickabaugh TM, Jamieson BD, Sun D, Li S, et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol*. 2016; 17:171. doi: 10.1186/s13059-016-1030-0
60. McGuire S. US Department of Agriculture and US Department of Health and Human Services, Dietary Guidelines for Americans, 2010. Washington, DC: US Government Printing Office, January 2011. *Advances in Nutrition: An International Review Journal*. 2011; 23:293-294.
61. Committee DGA. Scientific Report of the 2015 Dietary Guidelines Advisory Committee. Washington (DC):

- USDA and US Department of Health and Human Services. 2015.
62. Lu AT, Hannon E, Levine ME, Hao K, Crimmins EM, Lunnon K, Kozlenkov A, Mill J, Dracheva S, Horvath S. Genetic variants near MLST8 and DHX57 affect the epigenetic age of the cerebellum. *Nat Commun*. 2016; 7:10561. doi: 10.1038/ncomms10561
 63. Castelo-Branco C, Soveral I. The immune system and aging: a review. *Gynecol Endocrinol*. 2014; 30:16–22. doi: 10.3109/09513590.2013.852531
 64. Finch CE. Evolution in health and medicine Sackler colloquium: Evolution of the human lifespan and diseases of aging: roles of infection, inflammation, and nutrition. *Proc Natl Acad Sci USA*. 2010 (Suppl 1); 107:1718–24. doi: 10.1073/pnas.0909606106
 65. Chowdhury R, Stevens S, Gorman D, Pan A, Warnakula S, Chowdhury S, Ward H, Johnson L, Crowe F, Hu FB, Franco OH. Association between fish consumption, long chain omega 3 fatty acids, and risk of cerebrovascular disease: systematic review and meta-analysis. *BMJ*. 2012; 345:e6698. doi: 10.1136/bmj.e6698
 66. Zheng J-S, Hu X-J, Zhao Y-M, Yang J, Li D. Intake of fish and marine n-3 polyunsaturated fatty acids and risk of breast cancer: meta-analysis of data from 21 independent prospective cohort studies. *BMJ*. 2013; 346:f3706. doi: 10.1136/bmj.f3706
 67. Farina EK, Kiel DP, Roubenoff R, Schaefer EJ, Cupples LA, Tucker KL. Protective effects of fish intake and interactive effects of long-chain polyunsaturated fatty acid intakes on hip bone mineral density in older adults: the Framingham Osteoporosis Study. *Am J Clin Nutr*. 2011; 93:1142–51. doi: 10.3945/ajcn.110.005926
 68. Bannenberg GL, Chiang N, Ariel A, Arita M, Tjonahen E, Gotlinger KH, Hong S, Serhan CN. Molecular circuits of resolution: formation and actions of resolvins and protectins. *J Immunol*. 2005; 174:4345–55. doi: 10.4049/jimmunol.174.7.4345
 69. Ridker PM, Buring JE, Cook NR, Rifai N. C-reactive protein, the metabolic syndrome, and risk of incident cardiovascular events: an 8-year follow-up of 14 719 initially healthy American women. *Circulation*. 2003; 107:391–97. doi: 10.1161/01.CIR.0000055014.62083.05
 70. de Labry LO, Glynn RJ, Levenson MR, Hermos JA, LoCastro JS, Vokonas PS. Alcohol consumption and mortality in an American male population: recovering the U-shaped curve--findings from the normative Aging Study. *J Stud Alcohol*. 1992; 53:25–32. doi: 10.15288/jsa.1992.53.25
 71. Knott CS, Coombs N, Stamatakis E, Biddulph JP. All cause mortality and the case for age specific alcohol consumption guidelines: pooled analyses of up to 10 population based cohorts. *BMJ*. 2015; 350:h384. doi: 10.1136/bmj.h384
 72. Beach SR, Dogan MV, Lei MK, Cutrona CE, Gerrard M, Gibbons FX, Simons RL, Brody GH, Philibert RA. Methylopic Aging as a Window onto the Influence of Lifestyle: Tobacco and Alcohol Use Alter the Rate of Biological Aging. *J Am Geriatr Soc*. 2015; 63:2519–25. doi: 10.1111/jgs.13830
 73. Grønbaek M, Becker U, Johansen D, Gottschau A, Schnohr P, Hein HO, Jensen G, Sørensen TI. Type of alcohol consumed and mortality from all causes, coronary heart disease, and cancer. *Ann Intern Med*. 2000; 133:411–19. doi: 10.7326/0003-4819-133-6-200009190-00008
 74. Volpato S, Pahor M, Ferrucci L, Simonsick EM, Guralnik JM, Kritchevsky SB, Fellin R, Harris TB. Relationship of alcohol intake with inflammatory markers and plasminogen activator inhibitor-1 in well-functioning older adults: the Health, Aging, and Body Composition study. *Circulation*. 2004; 109:607–12. doi: 10.1161/01.CIR.0000109503.13955.00
 75. Fillmore KM, Kerr WC, Stockwell T, Chikritzhs T, Bostrom A. Moderate alcohol use and reduced mortality risk: systematic error in prospective studies. *Addict Res Theory*. 2006; 14:101–32. doi: 10.1080/16066350500497983
 76. Brien SE, Ronksley PE, Turner BJ, Mukamal KJ, Ghali WA. Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. *BMJ*. 2011; 342:d636. doi: 10.1136/bmj.d636
 77. Dauchet L, Amouyel P, Hercberg S, Dallongeville J. Fruit and vegetable consumption and risk of coronary heart disease: a meta-analysis of cohort studies. *J Nutr*. 2006; 136:2588–93.
 78. He FJ, Nowson CA, Lucas M, MacGregor GA. Increased consumption of fruit and vegetables is related to a reduced risk of coronary heart disease: meta-analysis of cohort studies. *J Hum Hypertens*. 2007; 21:717–28. doi: 10.1038/sj.jhh.1002212
 79. He FJ, Nowson CA, MacGregor GA. Fruit and vegetable consumption and stroke: meta-analysis of cohort studies. *Lancet*. 2006; 367:320–26. doi: 10.1016/S0140-6736(06)68069-0
 80. Carter P, Gray LJ, Troughton J, Khunti K, Davies MJ. Fruit and vegetable intake and incidence of type 2 diabetes mellitus: systematic review and meta-

- analysis. *BMJ*. 2010; 341:c4229. doi: 10.1136/bmj.c4229
81. Gandini S, Merzenich H, Robertson C, Boyle P. Meta-analysis of studies on breast cancer risk and diet: the role of fruit and vegetable consumption and the intake of associated micronutrients. *Eur J Cancer*. 2000; 36:636–46. doi: 10.1016/S0959-8049(00)00022-8
82. Wang X, Ouyang Y, Liu J, Zhu M, Zhao G, Bao W, Hu FB. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. *BMJ*. 2014; 349:g4490. doi: 10.1136/bmj.g4490
83. Holt EM, Steffen LM, Moran A, Basu S, Steinberger J, Ross JA, Hong C-P, Sinaiko AR. Fruit and vegetable consumption and its relation to markers of inflammation and oxidative stress in adolescents. *J Am Diet Assoc*. 2009; 109:414–21. doi: 10.1016/j.jada.2008.11.036
84. Esmailzadeh A, Kimiagar M, Mehrabi Y, Azadbakht L, Hu FB, Willett WC. Fruit and vegetable intakes, C-reactive protein, and the metabolic syndrome. *Am J Clin Nutr*. 2006; 84:1489–97.
85. Ornish D, Lin J, Chan JM, Epel E, Kemp C, Weidner G, Marlin R, Frenda SJ, Magbanua MJ, Daubenmier J, Estay I, Hills NK, Chainani-Wu N, et al. Effect of comprehensive lifestyle changes on telomerase activity and telomere length in men with biopsy-proven low-risk prostate cancer: 5-year follow-up of a descriptive pilot study. *Lancet Oncol*. 2013; 14:1112–20. doi: 10.1016/S1470-2045(13)70366-8
86. Farzaneh-Far R, Lin J, Epel ES, Harris WS, Blackburn EH, Whooley MA. Association of marine omega-3 fatty acid levels with telomeric aging in patients with coronary heart disease. *JAMA*. 2010; 303:250–57. doi: 10.1001/jama.2009.2008
87. Barzilai N, Crandall JP, Kritchevsky SB, Espeland MA. Metformin as a Tool to Target Aging. *Cell Metab*. 2016; 23:1060–65. doi: 10.1016/j.cmet.2016.05.011
88. Klempera P, Doubal S. A new approach to the concept and computation of biological age. *Mech Ageing Dev*. 2006; 127:240–48. doi: 10.1016/j.mad.2005.10.004
89. The Women’s Health Initiative Study Group. Design of the Women’s Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998; 19:61–109. doi: 10.1016/S0197-2456(97)00078-0
90. Patterson RE, Kristal AR, Tinker LF, Carter RA, Bolton MP, Agurs-Collins T. Measurement characteristics of the Women’s Health Initiative food frequency questionnaire. *Ann Epidemiol*. 1999; 9:178–87. doi: 10.1016/S1047-2797(98)00055-6
91. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012; 13:86. doi: 10.1186/1471-2105-13-86
92. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, Ritz BR, Chen B, Lu AT, Rickabaugh TM, Jamieson BD, Sun D, Li S, et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol*. 2016; 17:171. doi: 10.1186/s13059-016-1030-0
93. Whitsel EA, Baccarelli A, Hou L. Epigenetic Mechanisms of PM-Mediated CVD. Research Portfolio Online Reporting Tools. National Institutes of Health: U.S. Department of Health and Human Services: https://projectreporter.nih.gov/project_info_description.cfm?aid=8369149&icde=27485446.
94. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8:118–27. doi: 10.1093/biostatistics/kxj037
95. Pisani P, Faggiano F, Krogh V, Palli D, Vineis P, Berrino F. Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. *Int J Epidemiol*. 1997 (Suppl 1); 26:S152–60. doi: 10.1093/ije/26.suppl_1.S152
96. Bartali B, Frongillo EA, Bandinelli S, Lauretani F, Semba RD, Fried LP, Ferrucci L. Low nutrient intake is an essential component of frailty in older persons. *J Gerontol A Biol Sci Med Sci*. 2006; 61:589–93. doi: 10.1093/gerona/61.6.589
97. Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, Guralnik JM. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc*. 2000; 48:1618–25. doi: 10.1111/j.1532-5415.2000.tb03873.x

SUPPLEMENTARY INFORMATION

Selection of participants from the Women's Health Initiative BA23 dataset

Participants included individuals from the Women's Health Initiative (WHI), a national study that began in 1993 and enrolled postmenopausal women between the ages of 50-79 years into either randomized clinical trials (RCTs) or into an observational study; in our sample, 1,657 of these individuals were enrolled in at least one of the RCTs and the remaining 441 were enrolled in the observational study. For this analysis, women were selected from one of two large WHI sub-cohorts that had previously undergone genotyping as well as profiling for cardiovascular disease related blood biomarkers through two core WHI ancillary studies (Supplementary Figure 10).

The first cohort is known as the WHI SNP Health Association Resource (SHARe) cohort of minorities that includes >8000 African American women and >3500 Hispanic women. These women were genotyped through the WHI core study M5-SHARe (www.whi.org/researchers/data/WHIStudies/StudySites/M5) and underwent baseline blood biomarker profiling through WHI Core study W54-SHARe (www.whi.org/researchers/data/WHIStudies/StudySites/M5). The second cohort consists of a combination of European Americans from the two Hormonal Therapy (HT) trials selected for GWAS and biomarkers in core studies W58 (www.whi.org/researchers/data/WHIStudies/StudySites/W58) and W63 (www.whi.org/researchers/data/HIStudies/StudySites/W63).

While none of the participants of our study suffered from coronary heart disease at baseline, roughly half of the samples developed CHD at a later point. From the M5-SHARe and W54-SHARe cohorts, two sample sets were formed. The first (sample set 1) is a sample set of 637 CHD cases and 631 non-CHD cases as of Sept 30, 2010. The second sample set (sample set 2) is a non-overlapping sample of 432 cases of coronary heart disease and 472 non-cases as of September 17, 2012. We acknowledge a potential for selection bias using the above-described sampling scheme in WHI but suspect if such bias is present it is minimal. First, some selection bias is introduced by restricting our methylation profiling at baseline to women with GWAS & biomarker data from baseline as well, given the requirement that these participants must have signed the WHI supplemental consent for broad sharing of genetic data in 2005. However, we believe that selection bias at this stage is minimized by the inclusion of participants who died between the time of the start of the WHI study

and the time of supplemental consent in 2005, which resulted in the exclusion of only ~6-8% of all WHI participants. Nevertheless, participants unable or unwilling to sign consent in 2005 may not represent a random subset of all participants who survived to 2005. Second, some selection bias may also occur if similar gross differences exist in the characteristics of participants who consented to be followed in the two WHI extension studies beginning in 2005 and 2010 compared to non-participants at each stage. We believe these selection biases, if present, have minimal influence on our effect estimates.

A

		WHI BA23 dataset																							
		Adjusted for ethnicity				Caucasian				African American				Hispanic											
		n	μ	EAAA bicor	IEAA p	n	μ	EAAA bicor	IEAA p	n	μ	EAAA bicor	IEAA p	n	μ	EAAA bicor	IEAA p								
Diet	log2(Total energy)	2086	10.51	-0.02	0.33	0.00	0.87	992	10.58	-0.04	0.17	0.00	0.90	666	10.46	0.03	0.40	-0.01	0.85	428	10.45	-0.06	0.20	0.01	0.90
	Carbohydrate	2086	49.40	-0.01	0.59	0.04	0.05	992	49.45	0.01	0.81	0.05	0.11	666	49.03	-0.03	0.47	0.02	0.55	428	49.89	-0.02	0.61	0.06	0.19
	Protein	2086	16.54	-0.02	0.31	-0.03	0.24	992	16.89	-0.02	0.48	0.01	0.71	666	16.10	-0.02	0.70	-0.08	0.03	428	16.40	-0.02	0.65	-0.01	0.87
	Fat	2086	34.20	0.03	0.20	-0.02	0.45	992	33.23	0.01	0.71	-0.02	0.52	666	35.65	0.02	0.66	0.00	0.96	428	34.18	0.09	0.07	-0.05	0.35
	log2(1+Red meat)	2086	0.74	0.02	0.41	0.00	0.95	992	0.75	-0.02	0.51	-0.01	0.70	666	0.73	0.07	0.06	0.00	0.94	428	0.72	0.01	0.87	0.02	0.67
	log2(1+Poultry)	2086	0.46	-0.04	0.06	-0.05	0.02	992	0.41	-0.05	0.15	-0.05	0.09	666	0.52	-0.02	0.69	-0.08	0.04	428	0.47	-0.07	0.17	0.01	0.89
	log2(1+Fish)	2086	0.31	-0.08	6E-4	-0.02	0.34	992	0.29	-0.07	0.04	-0.02	0.52	666	0.37	-0.06	0.11	-0.09	0.03	428	0.25	-0.11	0.02	0.00	0.94
	log2(1+Dairy)	2086	1.25	-0.02	0.26	-0.02	0.43	992	1.42	-0.03	0.30	-0.01	0.72	666	1.02	0.01	0.84	-0.02	0.53	428	1.23	-0.06	0.24	-0.02	0.63
	log2(1+Whole grains)	2086	1.03	-0.01	0.58	0.01	0.76	992	1.00	-0.04	0.16	0.00	0.97	666	1.03	0.04	0.36	0.00	0.97	428	1.07	-0.02	0.63	0.03	0.52
	log2(1+Nuts)	2086	0.18	0.01	0.58	0.03	0.13	992	0.20	0.03	0.31	0.05	0.09	666	0.20	0.01	0.88	0.02	0.58	428	0.11	0.02	0.71	0.05	0.27
log2(Fruits)	2086	0.33	-0.03	0.20	0.02	0.40	992	0.47	-0.04	0.19	0.02	0.63	666	0.28	0.02	0.57	0.02	0.64	428	0.11	-0.07	0.12	0.02	0.65	
log2(Vegetables)	2086	0.60	-0.05	0.03	0.02	0.34	992	0.93	-0.06	0.08	0.02	0.59	666	0.44	0.01	0.85	0.01	0.89	428	0.10	-0.12	0.02	0.04	0.37	
Blood nutrients	Retinol	224	0.59	-0.09	0.18	0.00	0.99	74	0.65	0.05	0.70	0.21	0.08	85	0.57	-0.13	0.24	-0.10	0.38	65	0.56	-0.18	0.15	-0.15	0.24
	Mean carotenoids	224	0.03	-0.19	4E-3	-0.03	0.63	74	0.08	-0.21	0.07	-0.07	0.55	85	-0.02	-0.13	0.22	-0.04	0.75	65	0.03	-0.25	0.01	-0.01	0.95
	Lycopene	224	0.42	-0.12	0.07	-0.03	0.66	74	0.45	-0.11	0.36	0.07	0.56	85	0.37	-0.15	0.18	-0.09	0.40	65	0.44	-0.17	0.17	-0.09	0.49
	log2(alpha-Carotene)	224	-4.28	-0.16	0.02	0.01	0.88	74	-4.07	-0.14	0.24	-0.04	0.74	85	-4.60	-0.11	0.31	-0.02	0.85	65	-4.11	-0.22	0.07	0.05	0.70
	log2(beta-Carotene)	224	-2.09	-0.19	4E-3	-0.08	0.26	74	-1.94	-0.13	0.28	-0.03	0.77	85	-2.02	-0.16	0.14	-0.08	0.45	65	-2.35	-0.27	0.03	-0.11	0.38
	log2(Lutein+Zeaxanthin)	224	-2.36	-0.13	0.06	-0.06	0.38	74	-2.44	-0.08	0.52	-0.05	0.70	85	-2.22	-0.13	0.23	-0.16	0.15	65	-2.45	-0.17	0.17	0.08	0.54
	log2(beta-Cryptoxanthin)	224	-3.82	-0.04	0.52	0.03	0.67	74	-3.94	-0.15	0.19	-0.01	0.96	85	-3.82	-0.09	0.43	-0.08	0.48	65	-3.66	-0.06	0.66	0.02	0.88
	log2(alpha-Tocopherol)	224	3.90	-0.04	0.54	0.00	1.00	74	4.03	-0.10	0.40	0.03	0.78	85	3.81	0.04	0.75	-0.12	0.27	65	3.86	-0.05	0.69	0.13	0.30
	log2(gamma-Tocopherol)	224	0.87	0.05	0.48	0.12	0.07	74	0.87	0.00	0.98	0.07	0.54	85	0.96	0.15	0.17	-0.19	0.08	65	0.74	-0.01	0.93	0.08	0.55
	log2(C-reactive protein)	2073	1.58	0.11	3E-7	0.07	2E-3	991	1.26	0.12	3E-4	0.07	0.04	661	1.90	0.07	0.06	0.08	0.03	421	1.81	0.19	1E-4	0.05	0.33
log2(Insulin)	2051	5.82	0.09	3E-5	0.06	3E-3	970	5.65	0.12	2E-4	0.06	0.05	656	6.05	0.06	0.13	0.09	0.02	425	5.86	0.09	0.06	0.02	0.65	
log2(Glucose)	2091	6.70	0.04	0.08	0.06	0.01	995	6.67	0.07	0.02	0.05	0.15	666	6.74	0.02	0.62	0.07	0.07	430	6.70	0.00	0.93	0.07	0.17	
log2(Triglyceride)	2091	7.02	0.09	2E-5	0.08	2E-4	995	7.12	0.10	3E-3	0.08	0.02	666	6.76	0.08	0.04	0.09	0.02	430	7.21	0.11	0.02	0.08	0.09	
Total cholesterol	2091	233.03	0.04	0.06	0.03	0.16	995	236.31	0.08	0.01	0.04	0.24	666	231.75	0.02	0.66	0.01	0.79	430	227.37	0.00	0.98	0.04	0.37	
LDL cholesterol	2057	151.23	0.03	0.18	0.01	0.58	976	153.81	0.07	0.02	0.02	0.54	659	152.79	0.01	0.82	0.00	0.92	422	142.83	-0.03	0.50	0.02	0.74	
HDL cholesterol	2091	52.17	-0.06	4E-3	-0.04	0.09	995	51.01	-0.06	0.05	-0.06	0.06	666	54.71	-0.06	0.14	-0.02	0.62	430	50.91	-0.07	0.14	-0.02	0.64	
log2(Creatinine)	2041	-0.42	0.01	0.59	-0.01	0.67	973	-0.45	0.03	0.43	-0.02	0.62	657	-0.33	0.05	0.19	0.03	0.40	411	-0.50	-0.07	0.14	-0.07	0.19	
Systolic blood pressure	2093	132.02	0.05	0.01	0.02	0.26	995	131.92	0.10	2E-3	0.01	0.85	668	134.05	0.03	0.37	0.07	0.08	430	129.06	-0.02	0.63	-0.01	0.88	
Diastolic blood pressure	2093	76.41	0.03	0.24	0.03	0.13	995	75.21	0.06	0.05	0.02	0.49	668	78.71	0.00	0.98	0.08	0.05	430	75.58	-0.03	0.59	-0.01	0.78	
log2(Waist / hip ratio)	2093	-0.27	0.09	8E-5	0.04	0.10	995	-0.28	0.09	0.01	0.06	0.06	668	-0.27	0.06	0.13	-0.01	0.77	430	-0.28	0.13	0.01	0.06	0.23	
BMI	2093	29.85	0.08	6E-4	0.07	2E-3	995	28.83	0.10	1E-3	0.10	2E-3	668	31.71	0.05	0.20	0.06	0.13	430	29.32	0.06	0.20	0.01	0.81	
Socio-behavioral	Education	2071	6.66	-0.10	2E-6	0.00	0.83	989	6.86	-0.12	2E-4	0.02	0.43	660	6.85	-0.11	0.01	-0.03	0.40	422	5.90	-0.06	0.20	0.02	0.61
	Income	2019	3.54	-0.08	3E-4	-0.01	0.51	964	3.67	-0.06	0.06	0.02	0.64	642	3.47	-0.08	0.04	-0.06	0.11	413	3.32	-0.11	0.02	0.01	0.80
	log2(1+Exercise)	2090	2.49	-0.08	5E-4	-0.03	0.20	994	2.67	-0.08	0.01	0.01	0.70	667	2.30	-0.04	0.26	-0.11	4E-3	429	2.37	-0.12	0.01	0.02	0.67
	Current smoker	2090	0.14	0.00	0.93	0.00	0.88	994	0.13	-0.03	0.37	0.01	0.80	667	0.16	0.07	0.07	0.02	0.60	429	0.14	-0.01	0.77	0.05	0.27
	log2(1+Alcohol)	2086	1.09	-0.05	0.02	-0.01	0.52	992	1.44	-0.06	0.07	-0.02	0.48	666	0.73	-0.01	0.72	0.01	0.79	428	0.86	0.00	0.92	0.01	0.82

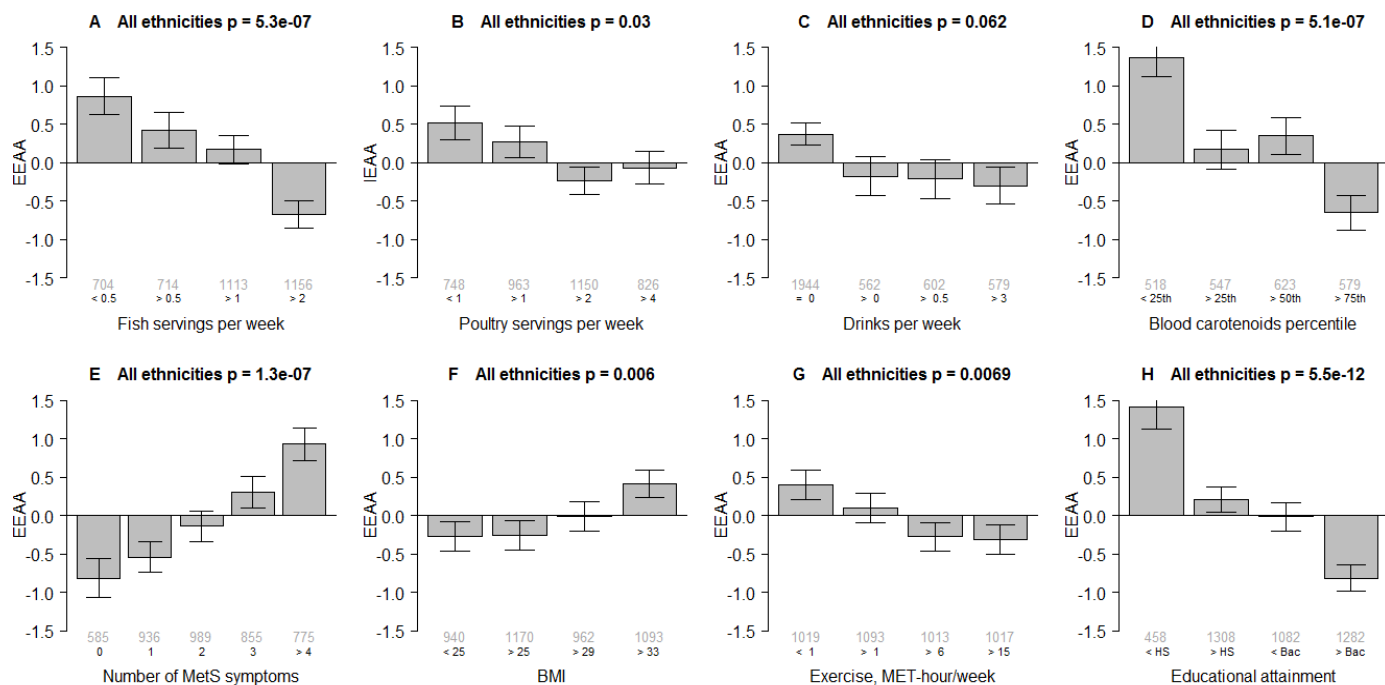
B

		WHI AS315 dataset																							
		Adjusted for ethnicity				Caucasian				African American				Hispanic											
		n	μ	EAAA bicor	IEAA p	n	μ	EAAA bicor	IEAA p	n	μ	EAAA bicor	IEAA p	n	μ	EAAA bicor	IEAA p								
Diet	log2(Total energy)	1601	10.56	-0.02	0.49	0.01	0.77	828	10.62	-0.03	0.37	0.05	0.13	405	10.49	0.04	0.45	0.03	0.52	205	10.58	-0.03	0.71	-0.14	0.05
	Carbohydrate	1601	48.50	0.02	0.47	-0.03	0.30	828	48.09	0.03	0.34	-0.04	0.24	405	48.67	0.03	0.49	0.01	0.88	205	48.34	-0.03	0.68	0.06	0.37
	Protein	1601	16.45	-0.03	0.27	-0.02	0.36	828	16.73	-0.06	0.07	-0.04	0.21	405	15.87	-0.05	0.29	-0.01	0.82	205	16.43	0.08	0.25	-0.01	0.87
	Fat	1601	35.28	0.01	0.58	0.03	0.23	828	34.98	0.03	0.37	0.06	0.11	405	36.32	-0.04	0.47	-0.03	0.57	205	35.65	0.05	0.44	-0.04	0.57
	log2(1+Red meat)	1601	0.77	0.02	0.42	0.07	0.01	828	0.79	0.02	0.59	0.10	3E-3	405	0.76	0.03	0.54	0.09	0.06	205	0.82	0.05	0.46	-0.09	0.20
	log2(1+Poultry)	1601	0.45	-0.02	0.43	-0.04	0.10	828	0.43	-0.05	0.19	-0.02	0.59	405	0.50	0.03	0.54	0.00	0.94	205	0.49	0.01	0.90	-0.16	0.02
	log2(1+Fish)	1601	0.32	-0.06	0.01	-0.01	0.59	828	0.31	-0.07	0.04	0.03	0.46	405	0.36	-0.05	0.27	-0.09	0.09	205	0.26	-0.03	0.72	-0.08	0.24
	log2(1+Dairy)	1601	1.26	-0.01	0.82	0.02	0.37	828	1.42	-0.04	0.27	0.04	0.27	405	1.00	0.05	0.27	0.08	0.11	205	1.30	0.06	0.37	-0.04	

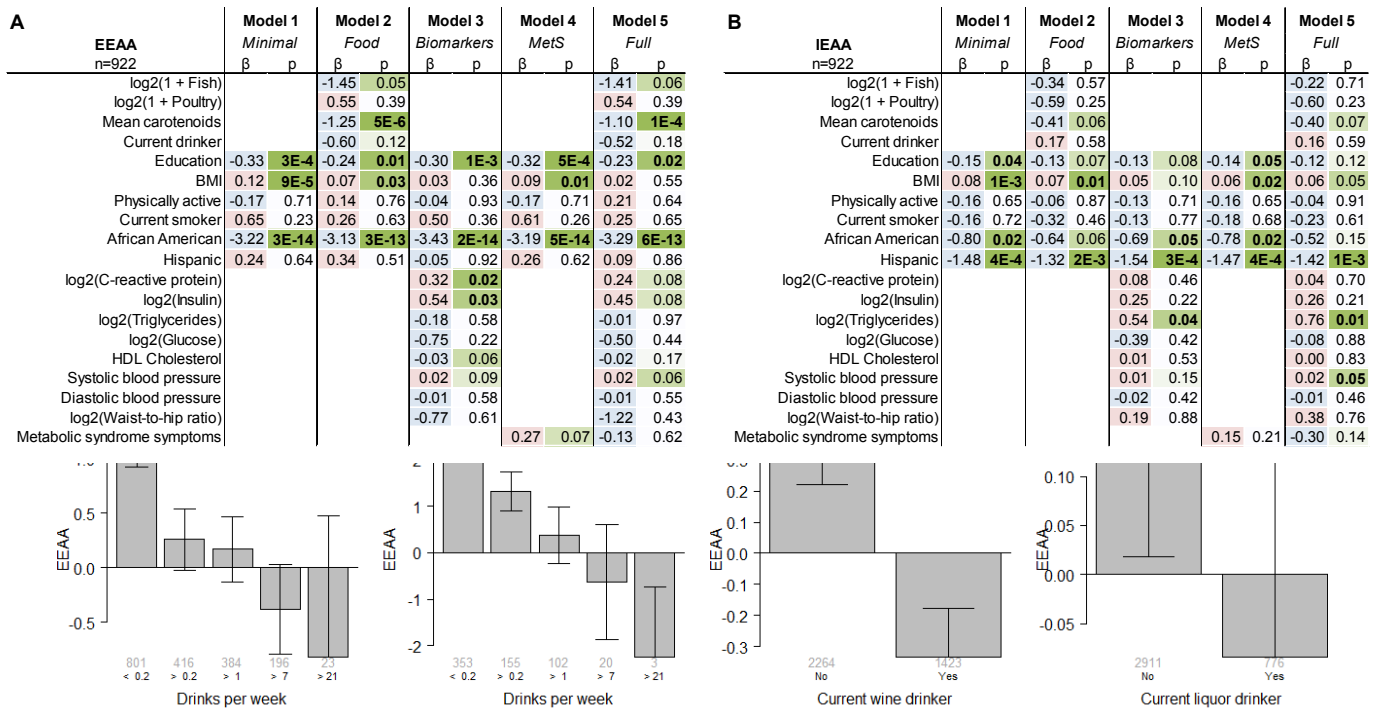
Supplementary Table 1. Linear models of IEAA and EEAA in WHI metformin users with repeat measurements.

	IEAA		EEAA	
	β Coefficient	P-Value	β Coefficient	P-Value
Started Metformin	-0.79	0.456	-0.003	0.998
IEAA/EEAA (First Blood Draw)	-0.33	2.88e-13	-0.35	8.92e-15
Age (First Blood Draw)	-0.55	1.73e-08	-0.96	4.67e-15
Age (Second Blood Draw)	0.50	6.02e-07	0.98	1.21e-14
Glucose (First Blood Draw)	0.02	0.132	0.03	0.048
Glucose (First Blood Draw)	0.01	0.657	-0.03	0.056
Education	0.27	0.004	-0.02	0.89
Race/Ethnicity (Reference=Non-Hispanic White)				
American Indian	-1.03	0.577	-1.31	0.555
Asian	-1.10	0.450	-3.54	0.044
Non-Hispanic Black	-2.04	0.166	-2.45	0.165
Hispanic	-0.63	0.658	-2.88	0.093
Other	1.32	0.614	-3.60	0.254

Table shows the beta-coefficients and corresponding p-values for linear models of IEAA and EEAA in the middle and right columns, respectively. The left column denotes the regressors in the models.



Supplementary Figure 2. EEAA among different levels of select dietary & lifestyle habits and measurements in the WHI. Panels A-H show bar plots visualizing the EEAA among stratified levels of fish, poultry (IEAA in this case), alcohol intake, blood carotenoid levels, number of metabolic syndrome symptoms, BMI, exercise, and education. Cut points roughly correspond roughly to quartiles except with number of MetS symptoms and alcohol intake which were selected for evenly-sized strata as much as possible. The sample sizes for each stratum are shown in grey beneath each bar. P-values for differences between strata are listed above each bar plot. Exercise is measured in units of metabolic equivalent hours per week and education uses high school diploma, some college, and bachelor's degree as cut points.



Supplementary Figure 4. EEAA among different strata of ethnic groups, levels and types of alcohol intake in the WHI. Panels A-D show bar plots visualizing the EEAA among stratified levels of alcohol intake (medium servings per week) for select ethnic groupings. Panels E-H show bar plots visualizing the EEAA among non- and current drinkers (at least one drink per month) of different types of alcoholic drinks: all types, beer, wine, and liquor. The sample sizes for each stratum are shown in grey beneath each bar. P-values for differences between strata are listed above each bar plot.

A

EEAA n	WHI								Meta-analysis	
	Caucasian 886		African 481		Hispanic 259		Asian 100		meta-t	meta-p
	β	p	β	p	β	p	β	p		
log2(1 + Fish)	-1.18	0.18	-0.99	0.28	-0.91	0.50	-0.67	0.73	-1.88	0.06
log2(1 + Poultry)	0.03	0.97	-0.03	0.98	0.47	0.65	1.82	0.35	0.41	0.68
Mean carotenoids	-0.98	2E-3	-0.93	0.02	-0.68	0.17	-1.25	0.16	-4.34	1E-5
Current drinker	-0.80	0.04	0.17	0.77	-0.95	0.18	-2.36	0.04	-2.31	0.02
Education	-0.05	0.62	-0.17	0.22	-0.47	2E-3	0.04	0.89	-2.20	0.03
BMI	0.03	0.43	0.14	2E-3	0.07	0.28	0.11	0.37	2.85	4E-3
Physically active	-0.54	0.28	0.47	0.44	1.06	0.19	-2.49	0.11	-0.24	0.81
Current smoker	1.15	0.06	0.75	0.35	-2.29	0.02	-0.82	0.76	0.87	0.38

B

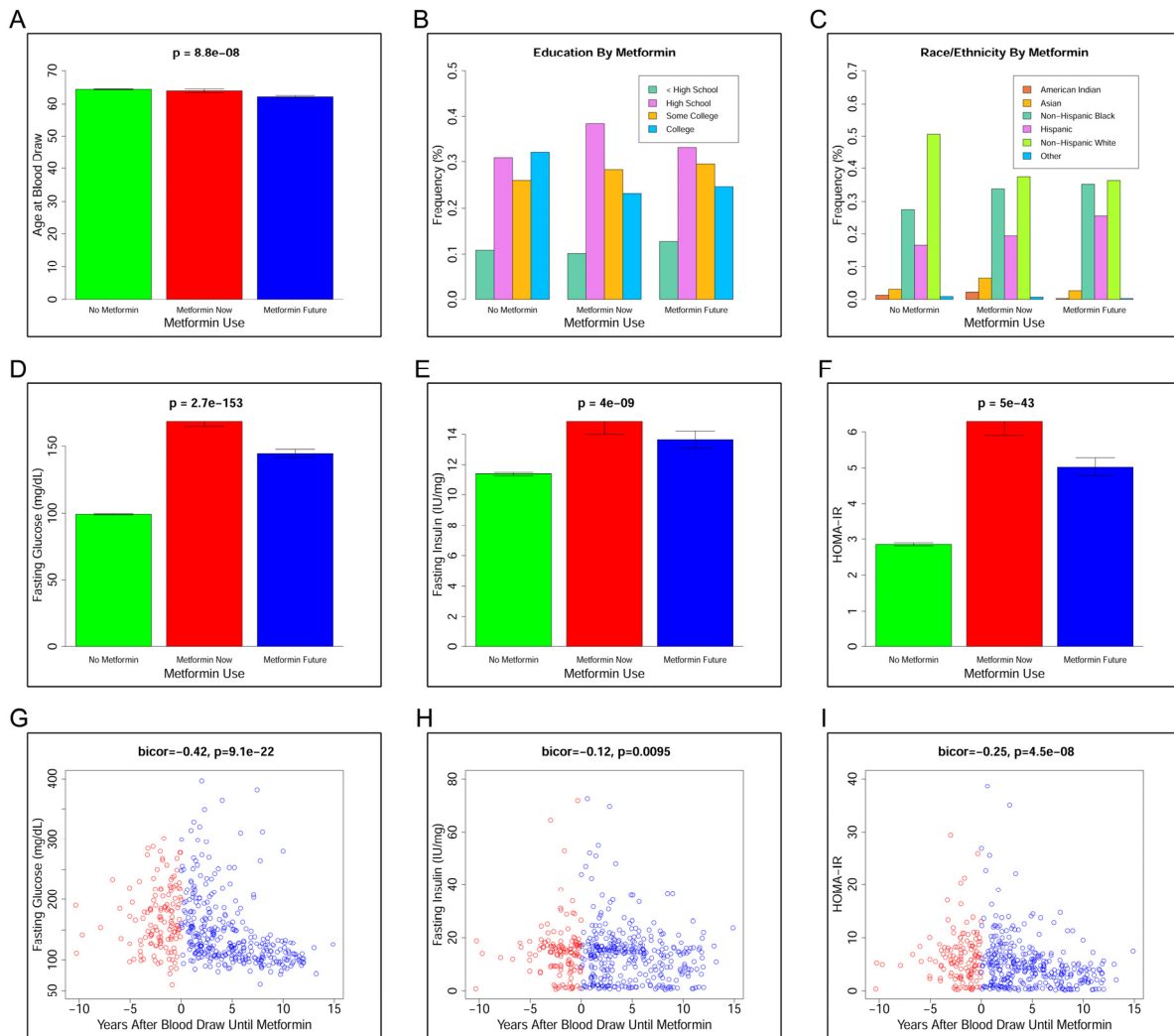
IEAA n	WHI								Meta-analysis	
	Caucasian 886		African 481		Hispanic 259		Asian 100		meta-t	meta-p
	β	p	β	p	β	p	β	p		
log2(1 + Fish)	0.20	0.78	-0.90	0.24	-0.21	0.85	1.05	0.52	-0.35	0.73
log2(1 + Poultry)	-0.50	0.43	0.13	0.85	-0.84	0.30	-2.23	0.18	-1.20	0.23
Mean carotenoids	-0.62	0.02	-0.43	0.20	-0.14	0.71	0.24	0.75	-2.47	0.01
Current drinker	-0.27	0.40	0.42	0.38	-0.42	0.46	1.09	0.27	-0.16	0.87
Education	-0.02	0.80	-0.22	0.06	-0.06	0.63	-0.35	0.15	-1.72	0.09
BMI	0.04	0.22	0.05	0.19	0.04	0.41	0.36	8E-4	2.72	0.01
Physically active	-0.10	0.82	0.05	0.92	0.91	0.16	-2.26	0.09	0.02	0.95
Current smoker	-0.17	0.74	-0.70	0.30	-0.55	0.48	0.13	0.96	-1.05	0.25

Supplementary Figure 5. Related to Figure 2. Meta-analysis of multivariable linear models of EEAA and IEAA including carotenoid levels in the WHI. Analogous to Figure 2 except including mean carotenoid levels: EEAA (panel A) and IEAA (panel B) were regressed on potential confounding factors, fish and poultry intake, mean across standardized measures of carotenoids, and current drinker status for the ethnic strata with sufficient sample sizes (n>100). Individual columns correspond to coefficient estimates (β) colored blue or red for negative and positive values respectively, and p-values (p) colored in green according to magnitude of significance, with the exception of the last two columns which denote Stouffer's method meta-t and meta-p values. Models are adjusted for originating dataset (WHI BA23 or WHI AS315).

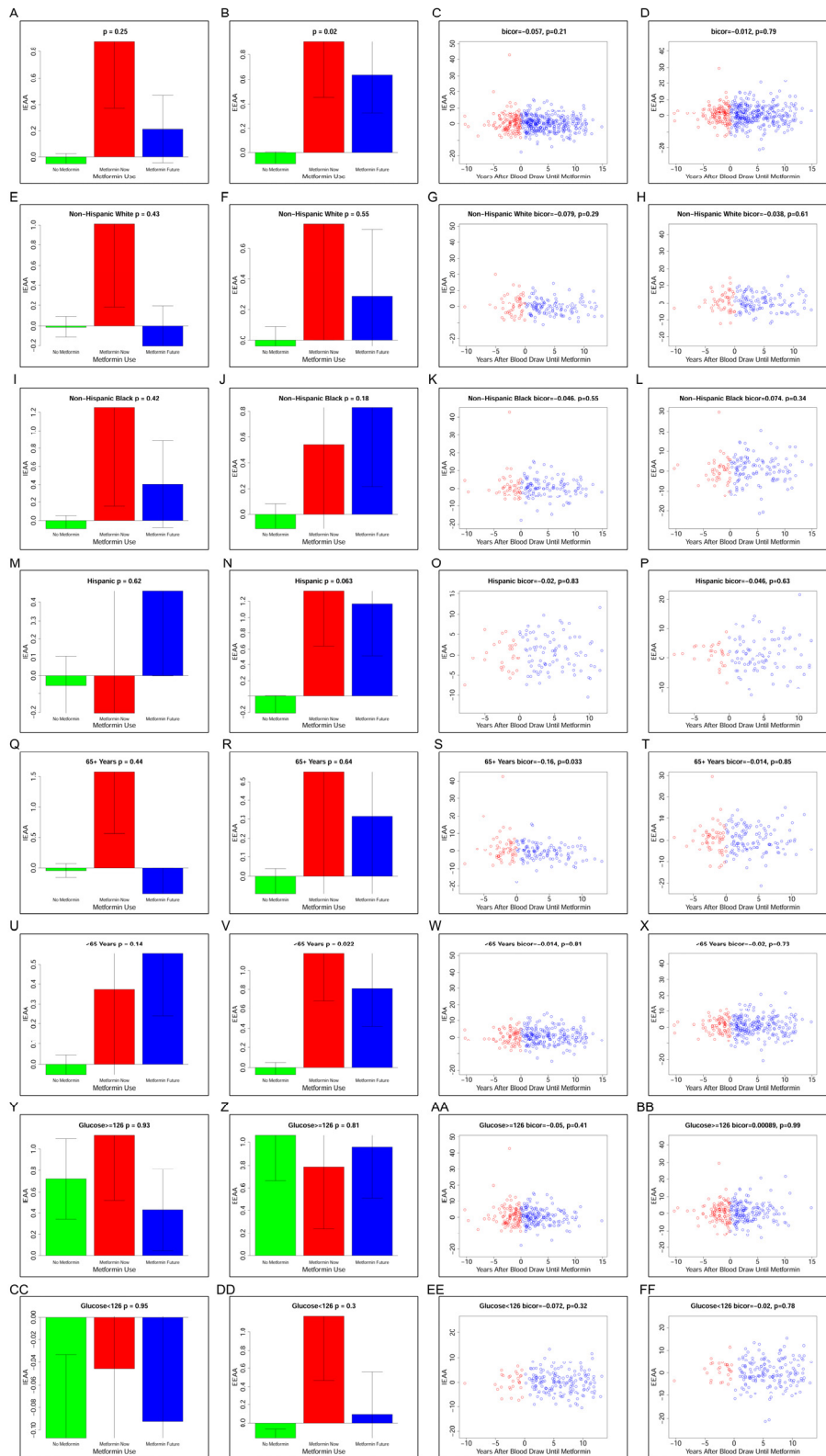
A	EEAA n=922	Model 1		Model 2		Model 3		Model 4		Model 5	
		Minimal		Food		Biomarkers		MetS		Full	
		β	p	β	p	β	p	β	p	β	p
	log2(1 + Fish)			-1.45	0.05					-1.41	0.06
	log2(1 + Poultry)			0.55	0.39					0.54	0.39
	Mean carotenoids			-1.25	5E-6					-1.10	1E-4
	Current drinker			-0.60	0.12					-0.52	0.18
	Education	-0.33	3E-4	-0.24	0.01	-0.30	1E-3	-0.32	5E-4	-0.23	0.02
	BMI	0.12	9E-5	0.07	0.03	0.03	0.36	0.09	0.01	0.02	0.55
	Physically active	-0.17	0.71	0.14	0.76	-0.04	0.93	-0.17	0.71	0.21	0.64
	Current smoker	0.65	0.23	0.26	0.63	0.50	0.36	0.61	0.26	0.25	0.65
	African American	-3.22	3E-14	-3.13	3E-13	-3.43	2E-14	-3.19	5E-14	-3.29	6E-13
	Hispanic	0.24	0.64	0.34	0.51	-0.05	0.92	0.26	0.62	0.09	0.86
	log2(C-reactive protein)					0.32	0.02			0.24	0.08
	log2(Insulin)					0.54	0.03			0.45	0.08
	log2(Triglycerides)					-0.18	0.58			-0.01	0.97
	log2(Glucose)					-0.75	0.22			-0.50	0.44
	HDL Cholesterol					-0.03	0.06			-0.02	0.17
	Systolic blood pressure					0.02	0.09			0.02	0.06
	Diastolic blood pressure					-0.01	0.58			-0.01	0.55
	log2(Waist-to-hip ratio)					-0.77	0.61			-1.22	0.43
	Metabolic syndrome symptoms							0.27	0.07	-0.13	0.62

B	IEAA n=922	Model 1		Model 2		Model 3		Model 4		Model 5	
		Minimal		Food		Biomarkers		MetS		Full	
		β	p	β	p	β	p	β	p	β	p
	log2(1 + Fish)			-0.34	0.57					-0.22	0.71
	log2(1 + Poultry)			-0.59	0.25					-0.60	0.23
	Mean carotenoids			-0.41	0.06					-0.40	0.07
	Current drinker			0.17	0.58					0.16	0.59
	Education	-0.15	0.04	-0.13	0.07	-0.13	0.08	-0.14	0.05	-0.12	0.12
	BMI	0.08	1E-3	0.07	0.01	0.05	0.10	0.06	0.02	0.06	0.05
	Physically active	-0.16	0.65	-0.06	0.87	-0.13	0.71	-0.16	0.65	-0.04	0.91
	Current smoker	-0.16	0.72	-0.32	0.46	-0.13	0.77	-0.18	0.68	-0.23	0.61
	African American	-0.80	0.02	-0.64	0.06	-0.69	0.05	-0.78	0.02	-0.52	0.15
	Hispanic	-1.48	4E-4	-1.32	2E-3	-1.54	3E-4	-1.47	4E-4	-1.42	1E-3
	log2(C-reactive protein)					0.08	0.46			0.04	0.70
	log2(Insulin)					0.25	0.22			0.26	0.21
	log2(Triglycerides)					0.54	0.04			0.76	0.01
	log2(Glucose)					-0.39	0.42			-0.08	0.88
	HDL Cholesterol					0.01	0.53			0.00	0.83
	Systolic blood pressure					0.01	0.15			0.02	0.05
	Diastolic blood pressure					-0.02	0.42			-0.01	0.46
	log2(Waist-to-hip ratio)					0.19	0.88			0.38	0.76
	Metabolic syndrome symptoms							0.15	0.21	-0.30	0.14

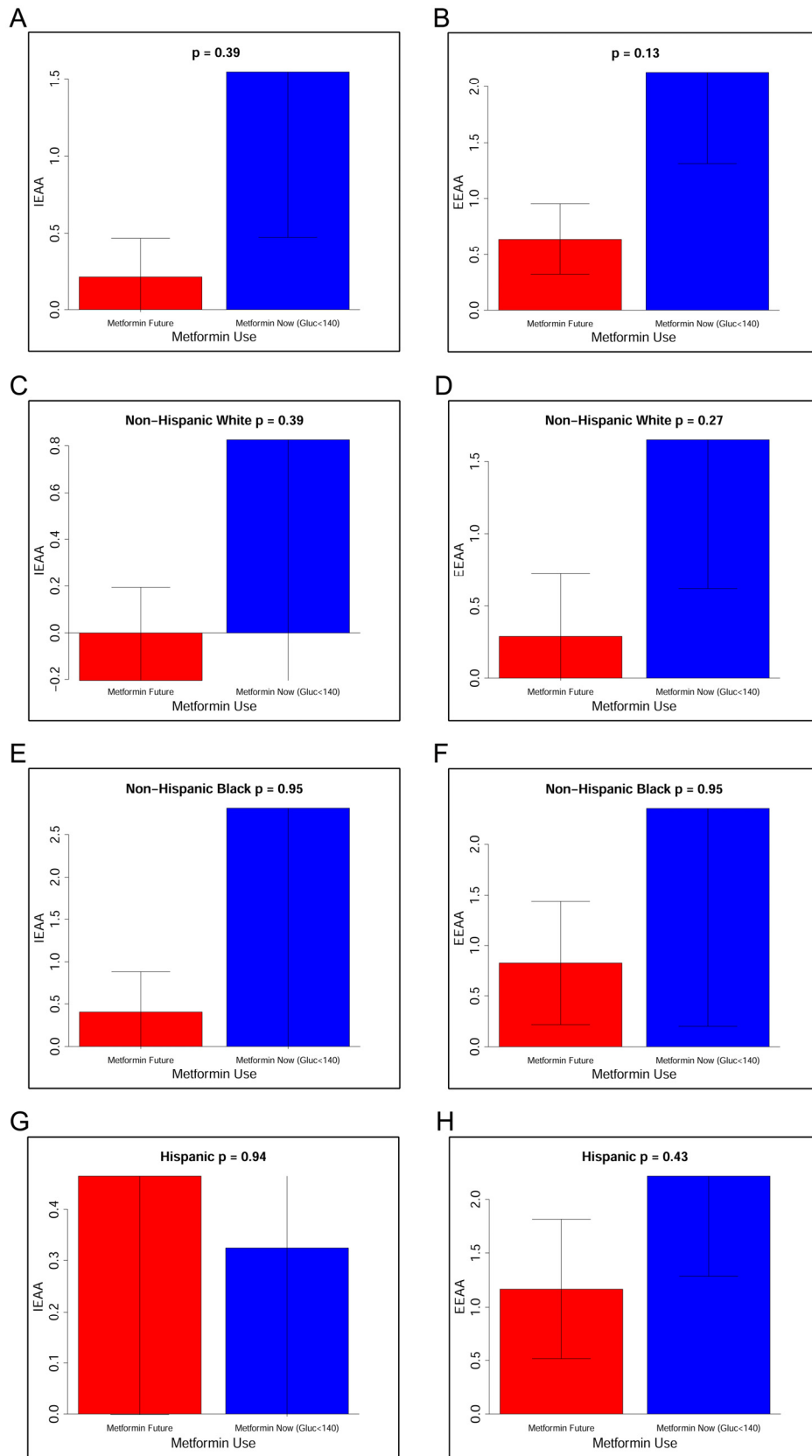
Supplementary Figure 6. Related to Figure 3. Multivariate linear models of EEAA and IEAA including carotenoid levels and with and without biomarkers in the WHI. Analogous to Figure 3 except including mean carotenoid levels: EEAA (panel A) and IEAA (panel B) were regressed on potential confounding factors, fish and poultry intake and current drinker status, and select biomarkers. Individual columns list the corresponding coefficient estimates (β) and p-values (p) for each fitting. Coefficients are colored according to sign (positive = red, negative = blue) and significance according to magnitude (green). Models 1 through 5 correspond to a minimal model, a model including dietary intake variables, a model including potential explanatory biomarkers, a model including number of metabolic syndrome symptoms and a complete model with all of the variables above, respectively. Models are adjusted for originating dataset (BA23 or AS315).



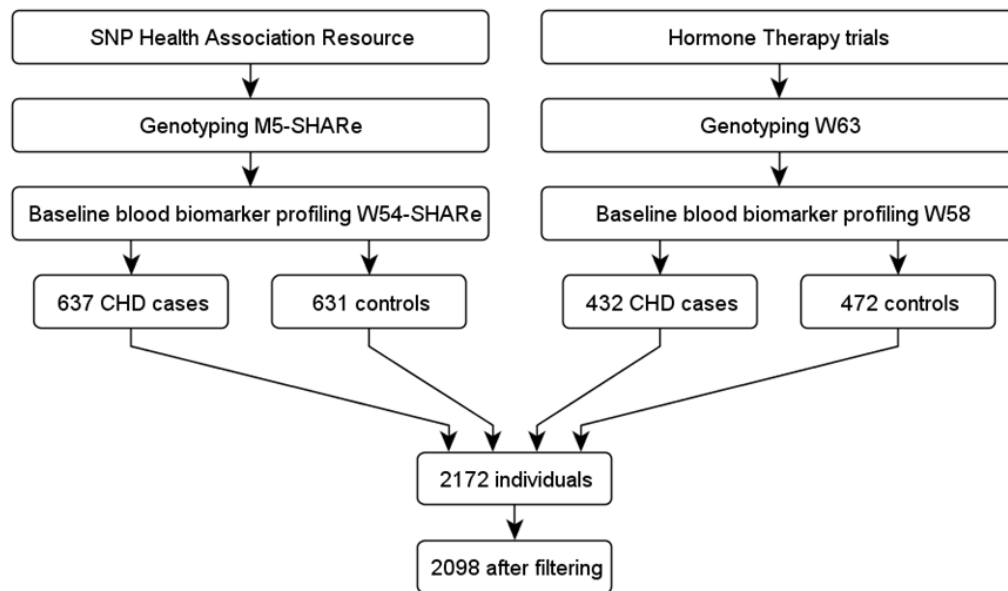
Supplementary Figure 7. Characteristics of non-, current, and future metformin users in the WHI. Panels A-F show barplots comparing the age at blood draw, educational attainment, ethnicity, fasting glucose, insulin, and insulin resistance of non-, current, and future metformin users (colored in green, red, and blue). Panels G-I show plotting fasting glucose, insulin, and insulin resistance as a function of years after blood draw until metformin with current and future users colored in red and blue respectively.



Supplementary Figure 8. IEAA and EAA among non-, current, and future users of metformin in the WHI. Columns 1 and 2 show barplots with IEAA and EAA on the y-axis and no metformin, current metformin, and later metformin usage strata (green, red, and blue respectively). Columns 3 and 4 show scatterplots with IEAA and EAA plotted as a function of years after blood draw until metformin usage with current and future metformin users colored in red and blue respectively. Rows show the same plots for different sociodemographic strata: all available WHI participants, non-hispanic white, non-hispanic black, hispanic, > 65 years old, < 65 years old, hyperglycemic, and non-hyperglycemic groups. P-values (and correlation coefficients) are shown and the top of each plot.



Supplementary Figure 9. IEAA and EEAA of future and non-hyperglycemic current metformin users in the WHI. Columns 1 and 2 show the IEAA and EEAA of future metformin users and current metformin users with fasting glucose levels of < 140mg/dL colored in red and blue respectively. Rows correspond to different ethnic strata: all ethnic groups, non-hispanic white, non-hispanic black, and hispanic. P-values for each comparison are shown at the top of each plot.



Supplementary Figure 10. Schematic denoting the selection of the WHI BA23 study sample. The study participants were originally selected for a case-control GWAS on coronary heart disease (CHD). Participants were selected from either the SNP Health Association Resource or the Hormone Therapy trials, and underwent genotyping and core blood biomarkers profiling. The data from these individuals were agglomerated and filtered for missing data.